DOCUMENT RESUME

ED 414 916 IR 056 774

TITLE	Scholarly Communication and Technology. Papers from the Conference Organized by the Andrew W. Mellon Foundation and Held at Emory University (Atlanta, Georgia, April 24-25, 1997).
PUB DATE	1997-04-00
NOTE	468p.; "A print publication, including the papers presented, a synthesis of the discussions, and some additional analysis of the topic will be made available at a later date by the University of California Press." For individual papers separately analyzed, see IR 056 775-799.
AVAILABLE FROM	The Web site of the Association of Research Libraries (ARL), which is hosting the papers electronically at: http://www.arl.org/scomm/scat/
PUB TYPE	Collected Works - Proceedings (021)
EDRS PRICE	MF01/PC19 Plus Postage.
DESCRIPTORS	Access to Information; Change; Computer Mediated Communication; *Conference Proceedings; Costs; *Electronic Journals; Electronic Publishing; Fair Use (Copyrights); Higher Education; *Information Technology; *Scholarly Journals; Standards; Users (Information)
IDENTIFIERS	Electronic Resources

ABSTRACT

This document includes 25 papers and conference summation remarks presented at the Scholarly Communication and Technology Conference. Issues under discussion during this 2-day event included the economics of electronic publishing, incorporating technology into academia, the future of consortia and access versus ownership, electronic content licensing, and updates on several electronic scholarly initiatives. Papers are divided according to the following nine sessions: (1) "The Economics of Electronic Publishing: Cost Issues"; (2) "The Evolution of Journals"; (3) "Economics of Electronic Publishing: Journals Pricing and User Acceptance"; (4) "Patterns of Usage"; (5) "Technical Choices and Standards"; (6) "Copyright and Fair Use"; (7) "Multi-Institutional Cooperation"; (8) "Sustaining Change"; (9) "Summation." (AEF)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

Scholarly Communication and Technology

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Richard Ekman

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)"

Papers from The Conference Organized by

The Andrew W. Mellon Foundation

at Emory University
April 24-25, 1997

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☐ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

ED 414 916

The Association of Research Libraries is pleased to host the web site for the papers presented at the conference, Scholarly Communication and Technology. The two-day event was organized by The Andrew W. Mellon Foundation and held at Emory University. It brought together a diverse group of people representing technologists, publishers, librarians, and scholars.

Issues under discussion during this two-day event included, the economics of electronic scholarly publishing, incorporating technology into academia, the future of consortia and access versus ownership, electronic content licensing, and updates on several electronic scholarly initiatives such as the Columbia University Online Books Project, Project Muse at Johns Hopkins University, and JSTOR.

A print publication, including the papers presented, a synthesis of the discussions, and some additional analysis of the topic will be made available at a later date by the University of California Press.

For additional information about the conference, or The Andrew W. Mellon Foundation's scholarly communication initiatives, please contact Richard Ekman. For additional information about ARL or this web site contact Patricia Brennan, ARL Program Officer at (202) 296-2296.

-ooooooo-

Copyright © of the papers on this site are held by the individual authors or The Andrew W. Mellon Foundation. Permission is granted to reproduce and distribute copies of these works for nonprofit educational or library purposes, provided that the author, source, and copyright notice are included on each copy. For commercial use, please contact **Richard Ekman** at the The Andrew W. Mellon Foundation (212) 838-8400.

IR056774

BEST COPY AVAILABLE

2

CONFERENCE PROGRAM

Thursday, April 24, 1997

Welcome

Billy E. Frye, Provost and Vice President for Academic Affairs, Emory University

Joan I. Gotwals, Vice-Provost and Director of Libraries, Emory University

Introductory Remarks

Richard Ekman, Secretary, The Andrew W. Mellon Foundation

Richard E. Quandt, Senior Advisor, The Andrew W. Mellon Foundation

Session #1 The Economics of Electronic Publishing: Cost Issues

Moderator: Richard Ekman

Comparing Electronic Journals to Print Journals: Are There Savings?
Janet Fisher, Associate Director, Journals Publishing, The MIT Press

Electronic Publishing in Academia: An Economic Perspective
Malcolm Getz, Associate Professor of Economics, Department of Economics and Business Administration, Vanderbilt University

Epic: Electronic Publishing is Cheaper
Willis G. Regier, Director, The Johns Hopkins University Press

The Use of Electronic Scholarly Journals Models of Analysis and Data Drawn from the Project Muse Experience at Johns Hopkins University
James G. Neal, Sheridan Director, Johns Hopkins University Library

The Library and the University Press: Two Views of the Costs and Problems of the Current System of Scholarly Publishing
Susan F. Rosenblatt, Deputy University Librarian, University of California at Berkeley

Economics of Electronic Publishing: Cost Issues - Comments on Session One Presentations
Robert Shirrell, Journals Manager, The University of Chicago Press

Session #2 The Evolution of Journals

Moderator: Richard E. Quandt

The Future of Electronic Journals
Hal Varian, Dean, School of Information, Management and Systems, University of California at Berkeley

Session #3 Economics of Electronic Publishing: Journals Pricing and User Acceptance

Moderator: Duane Webster, Executive Director, Association of Research Libraries

JSTOR: The Development of a Cost-Driven, Value-Based Pricing Model
Kevin M. Guthrie, Executive Director, JSTOR

The Effect of Price: Early Observations
Karen Hunter, Senior Vice President, Elsevier Science

The Economics of Electronic Journals
Andrew M. Odlyzko, Head, Mathematics and Cryptography Research Department, AT&T Research

Session #4 Patterns of Usage

Moderator: Gloria Werner, University Librarian, University of California at Los Angeles

Analysis of JSTOR: The Impact on Scholarly Practice of Access to On-line Journal Archives
Thomas A. Finholt, Assistant Professor of Psychology, Collaboratory for Research on Electronic Work, University of Michigan

Patterns of Use for the Bryn Mawr Reviews
Richard Hamilton, Paul Shorey Professor of Greek, Bryn Mawr College

Digital Libraries: A Unifying or Distributing Force?
Michael E. Lesk, Division Manager, Computer Science Research, Bellcore

Online Books at Columbia: Measurement and Early Results on Use, Satisfaction, and Effect
Carol A. Mandel, Deputy University Librarian, Columbia University, and
Mary C. Summerfield, Coordinator, Online Books Project, Columbia University Libraries

Cocktails and Dinner, Michael C. Carlos Museum

Host: Billy E. Frye, Provost and Vice President for Academic Affairs, Emory University

Digital Documents and the Future of the Academic Community
Dinner Speaker: Peter Lyman, University Librarian, University of California at Berkeley

....

Friday, April 25, 1997

Session #5 Technical Choices and Standards

Moderator: Ira Fuchs, Vice President for Computing and Information Technology, Princeton University

Making Technology Work for Scholarship: Investing in the Data
Susan Hockey, Department of English, University of Alberta

Technical Standards and Medieval Manuscripts
Brother Eric Hollas, OSB, Director, Hill Monastic Manuscript Library, Saint John's University

Digital Image Quality: From Conversion to Presentation and Beyond
Anne R. Kenney, Associate Director, Department of Preservation, Cornell University Library

Session #6 Copyright and Fair Use

Moderator: Jerry Campbell, University Librarian and Dean of University Libraries, University of Southern California

The HYPATIA Project (toward ASCAP for Academics)
Jane Ginsburg, Morton L. Janklow Professor of Literary and Artistic Property Law, Columbia University School of Law

The Transition to Electronic Content Licensing: The Institutional Context in 1997
Ann S. Okerson, Associate University Librarian, Yale University

Session #7 Multi-Institutional Cooperation

Moderator: Elaine Sloan, Vice President for Information Services, Columbia University

The Cross Currents of Technology Transfer: The Czech and Slovak Library Information Network
Andrew Lass, Project Manager, Czech and Slovak Library Information Network, Mount Holyoke College

Consortial Access Versus Ownership
Richard W. Meyer, Director of Libraries, Elizabeth Coates Maddux Library, Trinity University

A New Consortial Model for Building Digital Libraries
Raymond K. Neff, Vice President for Information Services, Case Western Reserve University

Session #8 Sustaining Change

Moderator: Sanford G. Thatcher, Director, Pennsylvania State University Press

Information-Based Productivity
Scott Bennett, University Librarian, Yale University

Cost and Value in Electronic Publishing
James J. O'Donnell, Professor of Classical Studies and Vice Provost (Interim), Information Systems and Computing, University of Pennsylvania

Session #9 Summation

Moderator: Richard Ekman

Daniel E. Atkins, Dean, School of Information and Library Studies, University of Michigan

Edward W. Barry, President, Oxford University Press

Deanna B. Marcum, President, Commission on Preservation and Access

Closing Comments

Richard E. Quandt

……

For additional information about the conference, or The Andrew W. Mellon Foundation's scholarly communication initiatives, please contact Richard Ekman. For additional information about ARL or this web site contact Patricia Brennan, ARL Program Officer at (202) 296-2296.

Return to Office of Scholarly Communication Home Page

ARL Home   ARL Scholarly Communication and Technology Home Page

© Association of Research Libraries, Washington, DC

Web Design by Angelo F. Cruz
Maintained by ARL Web Administrator
Last Modified: September 8, 1997

Scholarly Communication and Technology

Conference Organized by The Andrew W. Mellon Foundation

at Emory University
April 24-25, 1997

Copyright © of the papers on this site are held by the individual authors or The Andrew W. Mellon Foundation. Permission is granted to reproduce and distribute copies of these works for nonprofit educational or library purposes, provided that the author, source, and copyright notice are included on each copy. For commercial use, please contact Richard Ekman at the The Andrew W. Mellon Foundation.

Session #1 Economics of Electronic Publishing: Cost Issues

Comparing Electronic Journals to Print Journals: Are There Savings?

Janet H. Fisher
Associate Director for Journals Publishing
The MIT Press

. . . .

Comparing Electronic Journals to Print Journals: Are There Savings?

Three years ago the rhetoric of academics and librarians alike urged publishers to get on with it -- to move their publications from print to electronic formats. The relentless pressure on library budgets from annual increases of ten to twenty percent in serials prices made many look to electronic publication as the savior that would allow librarians to retain their role in the scholarly communication chain. Academics and university administrators were urged to start their own publications and take back ownership of their own research. The future role of the publisher was questioned: What did they do after all? Since so many scholars were now creating their own works on computer, why couldn't they just put them up on the Net? Who needs proofreading, copyediting, and design anymore? And since technology has made it possible for everyone to become a publisher, surely electronic publication would be cheaper than print.

There have been quite a few experiments in the last three years trying to answer some of the questions posed by the emergence of the Internet, but few have yielded hard numbers to date. Most have been focused on developing electronic versions of print products, and some of those will be discussed by others at this conference. MIT Press took a piece of the puzzle that we saw as important in the long run and within the capabilities of a university-based journal publisher with space and staff constraints. Many of our authors had been using e-mail, listserves, discussion groups, etc. for ten years or more, and we wanted to be visible on the Internet early.

We decided it was easier, cheaper, and less of a financial risk to try publishing a purely electronic journal rather than reengineering our production and delivery process for our print journals when we had so little feedback about what authors and customers really wanted. Starting with *Chicago Journal of Theoretical Computer Science* (CJTCS), which was announced in late 1994 and which began publication in June of 1995, we began publishing our first purely electronic journal. CJTCS, as well as *Journal of Functional and Logic Programming* (JFLP) and *Journal of Contemporary Neurology* (JCN), are published article-by-article. We ask subscribers to pay an annual subscription fee, but we have not yet installed elaborate mechanisms to ensure that only those who pay have access to the full text. *Studies in Nonlinear Dynamics and Econometrics* (SNDE), begun in 1996, is published quarterly in issues with the full text password protected. Another issue-based electronic journal -- *Videre: Journal of Computer Vision Research* -- will begin publishing this summer. You can view these publicatons at our web site (http://www-mitpress.mit.edu/).

The lack of one format for all material available in electronic format has been a problem for these electronic journals and our production staff. The publication format varies from journal to journal based on several criteria:

the format most often received from authors

the content of the material (particularly math, tables, special characters)

cost to implement

availability of appropriate browser technology

CJTCS and JFLP are published in LaTeX and PostScript, in addition to PDF (Adobe's Portable Document Format) which was added in 1997. JCN is published in PDF and HTML (Hypertext Markup Language, the language of the World Wide Web) because the PostScript files were too large to be practical. SNDE is published in PostScript and PDF. Videre will be published in PDF.

Here I will be presenting our preliminary results on the costs of electronic only journals and comparing them to the costs of traditional print journals. I will be using *Chicago Journal of Theoretical Computer Science* as the model but will include relevant information from our experience with our other electronic journals.

Background on the Project

CJTCS was announced in fall of 1994 and began publication in June of 1995. Material is forwarded to us from the editor once the review process and revisions have been completed. Four articles were published from June through December of 1995, and six articles were published in 1996. (See appendix 1 for list of articles published.) The web site is hosted at the University of Chicago, with entry from the MIT Press web site. The production process includes the following steps:

copyediting

return of copyedited manuscript to author

author's response goes back to copyeditor

final copyedited article goes to "typesetter"

typesetter enters edits/tagging/formatting

proofreading

author sees formatted version

typesetter makes final corrections

article is published (i.e., posted on the site)

Tagging and "typesetting" has been done by Michael J. O'Donnell, Managing Editor of CJTCS who is a professor at University of Chicago.

The subscription price is $30/year for individuals and $125/year for institutions. When an article is published, subscribers receive an e-mail message announcing its publication. Included is the title, the author, the abstract, the location of the file, and the articles published to date in the volume. Articles are numbered sequentially in the volume (e.g., 1996-1, 1996-2). Individuals and institutions are allowed to use the content liberally, with permission to do the following posted on the Web site:

- read articles directly from the official journal servers, or from any other server that grants you access
- copy articles to your own file space for temporary use
- form your own permanent archive of articles, which you may keep even after your subscription lapses
- display articles in the ways most convenient to you (on your computer, printed on paper, converted to spoken form, etc.)
- apply agreeable typographical styles from any source to lay out and display articles

- apply any information retrieval, information processing, and browsing software from any source to aid your study of articles
- convert articles to other formats from the LaTeX and PostScript forms on the official servers
- share copies of articles with other subscribers
- share copies of articles with nonsubscribing collaborators as a direct part of your collaborative study or research

Library subscribers may also:

- print individual articles and other items for inclusion in your periodoical collection or for placing on reserve at the request of a faculty member
- place articles on your campus network for access by local users, or post article listings and notices on the network
- share print or electronic copy of articles with other libraries under standard interlibrary loan procedures

In February 1996, Michael O'Donnell installed a HyperNews feature to accompany each article which allows readers to give feedback on articles. Forward pointers, which were planned to update the articles with appropriate citations to other material published later, have not yet been instituted. Although the editors originally envisioned these features as very important to readers, no questions or comments about the articles have been posted to date.

Archiving arrangements were made with (1) the MIT Libraries, which is creating archival microfiche and archiving the PostScript form of the files; (2) MIT Information Systems, which is storing the LaTeX source on magnetic tape and refreshening it periodically; and (3) the Virginia Polytechnic Institute Scholarly Communications Project, which is mirroring the site (http://scholar.lib.vt.edu/).

Direct Costs of Publication

To date, CJTCS has published ten articles with a total of 244 pages. I have chosen to compare the direct costs we have incurred in publishing those 244 pages with the direct costs we incurred for a 244-page issue (Volume 8, Number 5, July 1996) of another of our journals, *Neural Computation* (NC). NC has a print run of approximately 2000 copies, and typesetting is done from LaTeX files supplied by the authors (as is the case for CJTCS):

	CJTCS	NC	% Difference
Copyediting/Proofreading	$ 1,114	$ 1,577	+ 42%
Composition	$ 2,070	$ 3,914	+ 89%
Printing & Binding	--	$ 6,965	---
Total Production Cost	$ 3,184	$12,456	+291%

Composition Cost/Page	\$ 8.48	\$ 16.24	+ 92%
Total Production Cost/Page	\$ 13.05	\$ 51.05	+291%

Important differences in production processes that affect these costs are:

1. The number of articles published (ten in CJTCS, 12 in NC).
2. The copyeditor handles author queries for NC and bills us hourly. This contributed \$100 to the copyediting bill.
3. Composition for CJTCS is done on a flat fee basis of \$200. Tagging and formatting has been done by Michael O'Donnell, the journal's Managing Editor at University of Chicago, because we were unable to find a traditional vendor willing to tag on the basis of content rather than format. The \$200 figure was developed in conjunction with a LaTeX coding house we planned to use initially but which was unable to meet the journal's schedule requirements. In comparison, the cost per article for NC is approximately \$326, which includes a \$58/article charge for producing repro pages to send to the printer and a \$21/article charge for author alteration charges. These are not included on the CJTCS composition bills.

The overhead costs associated with CJTCS and this issue of *Neural Computation* vary greatly. Overhead for our print journals is allocated on the following basis:

- Production -- charged to each journal based on the number of issues published
- Circulation -- charged to each journal based on the number of subscribers, the number of issues published, whether the journal has staggered or non-staggered renewals, and whether copies are sold to bookstores and newsstands
- Marketing/General and Administrative -- divided evenly among all journals.

For CJTCS, the Press incurs additional overhead costs associated with the Digital Projects Lab (DPL). These include the cost of staff, and hardware and software associated with the Press's World Wide Web server. These are allocated to each electronic publication on the following basis:

- Cost of hardware and software for the fileserver, network drops, staff time spent maintaining the server, etc., allocated to each e-journal based on the percentage of disk space the journal files occupy as a function of all web-related files on our server
- Amount of time per issue or article that DPL staff work on the journal times the rate per hour of staff

A comparison of overhead costs associated with CJTCS and this issue of *Neural Computation* shows:

	CJTCS	NC 8:5

Journals Department		
Production	$ 8,000	$ 1,000
Fulfillment Cost Per Subscriber	$ 108	1
General and Administrative	$ 31,050	$ 2,300
Digital Projects Lab		
Staff	$ 200	--
Hardware and Software	$ 5,000	--
Total Overhead Per Subscriber	$ 44,358	$ 3,301
OH costs per page published	$ 182	$ 14

For comparison, below are the direct costs associated with three other electronic journals to date: *Journal of Contemporary Neurology* (JCN), *Journal of Functional and Logic Programming* (JFLP), and *Studies in Nonlinear Dynamics and Econometrics* (SNDE).

	# Pages	# Articles/Issues	Direct Costs	Cost/Pg
JCN	34	6 articles	$ 1666	$ 49.00
JFLP	280	7 articles	$ 2204	$ 7.87
SNDE	152	2 issues	$ 4184	$ 27.53

JCN's cost/page is much higher than the other e-journals because the typesetter produces PDF and HTML formats and deals wtih complex images. It also takes additional time from our Digital Projects Lab staff because of the HTML coding and linking of illustrations, which adds an additional $7.00 per page to its costs. The total cost per page for JCN is, therefore, in line with our print journals even though there is no printing and binding expense.

The issue-based electronic journal *Studies in Nonlinear Dynamics and Econometrics* (SNDE) is comparable in direct costs with a standard print journal, with the only difference being the lack of printing and binding costs. Below is a comparison of the direct costs incurred for SNDE 1:1 (76 pages) and an 80-page issue of one of our print journals, *Computing Systems* (COSY), that follows a similar production path:

	SNDE 1:1	COSY 8:4
Copyediting/Proofreading	$ 551	$ 554
Composition	$ 1,383	$ 1,371
Printing and Binding	$ --	$ 6,501
Total Production Cost	$ 1,934	$ 8,426
Comp Per Page	$ 18.20	$ 17.57

```
     Total Production Cost Per Page      $   25.44        $ 105.33
```

Composition cost per page is comparable in these journals, but the total production cost per page of SNDE is only 24% of that of COSY which includes the printing and binding costs associated with a 6000-copy print run. The overhead costs, however, are higher for the electronic journal because of the addition of $1,400 per issue in indirect costs incurred for the staff, hardware, and software in the Digital Projects Lab.

Market Differences

The other side of the picture is whether the market reacts similarly to electronic only products. Since this question is outside the scope of this paper, I will only generalize here from our experience to date. For the four electronic journals we have started, the average paid circulation to date is approximately 100, with 20 to 40 of those being institutional subscriptions. For the two print journals we started in 1996 (both in the social sciences), the average circulation at the end of their first volumes (1996) was 550, with an average of 475 individuals and 75 institutions. There appears to be a substantial difference in the readiness of the market to accept electronic only journals at this point, as well as reluctance on the part of the author community to submit material. It is, therefore, more difficult for the publisher to reach break even with only one-fifth of the market willing to purchase, unless subscription prices are increased substantially. Doing this would likely dampen the paid subscriptions even more.

Conclusion

From the comparison between CJTCS and *Neural Computation*, it seems that the direct costs of publishing an electronic journal are substantially below that of a print journal with comparable pages. The overhead costs, however, are much higher -- 1240% higher in this case -- but that is adversely affected by the small amount of content published in CJTCS over the course of 18 months of overhead costs compared with NC which published 12 issues over the same period of time. The disparity in the markets for electronic products and print products is, at this point in time, a very big obstacle to their financial viability, as is also the conservatism of the author community.

....

For additional information about the conference, or The Andrew W. Mellon Foundation's scholarly communication initiatives, please contact Richard Ekman. For additional information about ARL or this web site contact Patricia Brennan, ARL Program Officer at (202) 296-2296.

Return to Office of Scholarly Communication Home Page

ARL Home ARL Scholarly Communication and Technology Home Page

© Association of Research Libraries, Washington, DC
Web Design by Angelo F. Cruz
Maintained by ARL Web Administrator
Last Modified: August 29, 1997

Scholarly Communication and Technology

Conference Organized by The Andrew W. Mellon Foundation

at Emory University
April 24-25, 1997

Copyright © of the papers on this site are held by the individual authors or The Andrew W. Mellon Foundation. Permission is granted to reproduce and distribute copies of these works for nonprofit educational or library purposes, provided that the author, source, and copyright notice are included on each copy. For commercial use, please contact Richard Ekman at the The Andrew W. Mellon Foundation.

Session #1 Economics of Electronic Publishing: Cost Issues

Electronic Publishing in Academia: An Economic Perspective

Malcolm Getz
Associate Professor of Economics
Department of Economics and Business Administration
Vanderbilt University

……

Malcolm Getz
May 21, 1997
Malcoml.Getz@Vanderbilt.edu
http://www.vanderbilt.edu/Econ/MGetz.html

Electronic Publishing in Academia

An Economic Perspective[1]

The Library at Washington University reports 150,000 hits per year on its electronic, networked *Encyclopedia Britannica* at a cost to the Library of four cents per hit.[2] This rate of use seems to be an order of magnitude larger than the rate of use of the print version of the document in the library. At the same time, the volunteer Project Gutenberg whose goal was to build an electronic file of 10,000 classic, public domain texts on the Internet has failed to sustain itself.[3] The University of Illinois decided it could no longer afford to provide the electronic storage space and no other entity stepped forward to sustain the venture.

A first lesson here is that production values, the quality of indexing and presentation, the packaging and marketing of the work, matter. Those ventures that take the approach of unrestricted free access don't necessarily dominate ventures that collect revenues. When a shopper asks "What does it cost?" we can naturally respond "What is it worth to you?" Electronic communication among academics is growing when it is valuable. In contemplating investments in electronic publishing, the publisher's, and indeed academia's, goal is to create the most value for the funds invested. Generally, the freebie culture that launched the Internet represents only a subset of a much wider range of possible uses. Many quality information products that flow through the Net will be generating revenue flows sufficient to sustain them.

The *Encyclopedia* gives a second lesson, namely, that the costs of electronic distribution may be significantly less than print. Serviceable home encyclopedias on CD now cost about $50 and *Britannica* is about $300, a small fraction of the price of the print editions of the same encyclopedias just a few years ago. Indeed, the latest word processing software includes tools that will allow anyone who uses word processing to create documents tagged for posting on the World Wide Web. Essentially, anyone who owns a current vintage computer with sufficient network connection can make formatted text with tables and graphics available instantly to everyone on the Net. The cost of such communication is a small fraction of the cost of photocopying and mailing documents.

An important consequence of the dramatic decline in the cost of sharing documents is the likelihood of a dramatic increase in the quantity of material available. Everyone who writes may post the whole history of their work on the web at little incremental cost. Availability is then hardly an issue.

The challenge to academia is to invest in services that will turn the ocean of data into sound, useful, compelling information products. The process of filtering, labeling, refining, and packaging, that is, the process of editing and publishing, takes resources and will be shaped by the electronic world in significant ways. This essay is concerned with this process.

Scholar

Begin with first principles. Academia may become more useful to our society at large by communicating electronically. When electronic scholarship is more valuable, our institutions will invest more.

Scholarship plays three roles in our society. First, academia educates the next generation of professionals, managers, and leaders. Second, it makes formal knowledge available to society at large, stimulating the development of new products, informing debates on public policy, and improving understanding of our culture. Third, it develops new knowledge. Digital communication ought ultimately to be judged by how well it serves these three activities, teaching, service, and research. Consider each in turn.

Access to networked, digital information is already enhancing education. More students at more institutions have access to more information because of the World Wide Web. About 60 percent of high school graduates now pursue some college, and President Clinton has called for universal access to two years of college.[4] The importance of the educational mission is growing. Of course, today networked information is sporadic and poorly organized relative to what it might someday become. Still, the available search services, rapid access, and the wide availability of the network are sufficient to demonstrate the power of the tool. Contrast the service with a conventional two-year college library whose size depends on the budget of the institution, when access often depends on personal interaction with a librarian, and where a student must plan a visit and sometimes even queue for service. Access to well-designed and supported Web-based information gives promise of promoting a more active style of education. Students may have more success with more open-ended assignments, participate in on-line discussion with others pursuing similar topics, and get faster feedback from more colorful, more interactive materials. Integrating academic information into the wider universe of Web information seems likely to have important benefits for students when it is done well.

Similarly, many audiences for academic information outside the walls of the academy already use the World Wide Web. Engineering Information, Inc., (EI) for example, maintains a subscription web site for both academic and non-academic engineers.[5] A core feature of the service is access to the premier index to the academic engineering literature with a fulfillment service. But EI's Village offers on-line access to professional advisers, conversations with authors, and services for practicing engineers. Higher quality, more immediate access to academic information seems likely to play an increasing role in the information sectors of our society, including nearly every career where some college is a common prerequisite. Higher education seems likely to find wider audiences by moving its best materials to the networked, digital arena.

In the business of generating new knowledge, the use of networked information is already accelerating the pace. Working papers in physics, for example, are more rapidly and widely accessible from the automated posting service at Los Alamos than could possibly be achieved by print.[6] In text oriented fields, scholars are able to build concordances and find patterns in ways impossible with print. Duke University's digital papyrus, for example, offers images of papyri with rich, searchable descriptive information in text.[7] In economics, the web gives the possibility of mounting data sets and algorithmic information and so allows scholars to interact with the work of others at a deeper level than is possible in print. For example, Ray Fair maintains his 130 equation model of the US economy on the web with data sets and a solution method.[8] Any scholar who wants to experiment with alternative estimations and forecasting assumptions in a fully developed simulation model may do so with modest effort. In biology, the Human Genome Project is only feasible because of the ease of electronic communication, the sharing of databases, and other on-line tools.[9] In visually oriented fields, digital communication offers substantial benefits, as video and sound may be embedded in digital documents. Animated graphics with sound may have significant value in simulation models in science. In art and drama, digital files may allow comparative studies previously unimaginable. Digital communication, then, may have its most significant consequence in accelerating the development of new knowledge.

The pace of investment in digital communication within academia may well be led by its value in education, service broadly defined, and research. In each case, institutional revenues and success may depend on effective deployment of appropriate digital communication. Of course, individual scholars face a significant challenge in mastering the new tools and employing them in appropriate ways. It is also worth emphasizing that not all things digital are valuable. However, when digital tools are well used, they are often significantly more valuable than print.

Publisher

The evolution of the digital arena will be strongly influenced by cost and by pricing policies. Cost is a always a two-way street, a reflection, on the one hand, of the choices of authors and publishers who commit resources to publication and, on the other, of the choices of readers and libraries who perceive value. Publishers are challenged to harvest raw materials from the digital ocean and fashion valuable information products. Universities and their libraries must evaluate the possible ways of using digital materials and restructure budgets to deploy their limited resources to best advantage. Between publisher and library stands the electronic agent who may broker the exchange in new ways. Consider first the publisher.

The opportunity to distribute journals electronically has implications for the publishers' costs and revenues. On the cost side, the digital documents can be distributed at lower cost than paper. The network may also reduce some editorial costs. However, sustaining high production values will continue to involve considerable cost because quality editing and presentation are costly. On the revenue side, sale of individual subscriptions may, to some degree, yield to licenses for access via campus intranets and to pay-per-look services.

Publisher Costs

The central fact of the publishing business is the presence of substantial fixed cost with modest variable cost. The cost of gathering, filtering, refining, and packaging shapes the quality of the publication but does not relate to distribution. The cost of copying and distributing the publication is a modest share of the total expense. A publication with high production values will have high fixed costs. Of course, with larger sale, the fixed costs are spread more widely. Thus, popular publications have lower cost per copy because each copy need carry only a bit of the fixed cost. In thinking about a digital product, the publisher is concerned to invest sufficiently in fixed costs to generate a readership that will pay prices that cover the total cost.

There is a continuum of publications, from widely distributed products with high fixed costs but lower prices to narrowly distributed products with low fixed costs but higher prices. We might expect an even wider range of products in the digital arena.

To understand one end of the publishing spectrum, consider a publisher who reports full financial accounts and is willing to share internal financial records, namely, the American Economic Association (AEA). The AEA is headquartered in Nashville but maintains editorial offices for each of its three major journals in other locations. The AEA has 21,000 members plus 5,500 additional journal subscribers. Membership costs between $52 and $73 per year (students $26) and members get all three journals. The library rate is $140 per year for the bundle of three journals. The Association had revenues and expenditures of $3.7 million in 1995.

The AEA prints and distributes nearly 29,000 copies of the *American Economic Review* (AER), the premier journal in economics. The AER receives nearly 900 manuscripts per year and publishes about 90 of them in quarterly issues. A *Papers and Proceeding* issue adds another 80 or so papers from the Association's annual meeting. The second journal, the *Journal of Economic Perspectives* (JEP) invites authors to contribute essays and publishes more topical, less technical essays, with 56 essays in four issues in 1995. The third journal, the *Journal of Economic Literature* (JEL) contains an index to the literature in economics, indexing and abstracting several hundred journals, listing all new English-language books in economics, and reviewing nearly 200 books per year. The JEL publishes more than 20 review essays each year in four quarterly issues. The three journals together yield about 5,000 pages, about 10 inches of linear shelf space, per year. The index to the economic literature published in JEL is cumulated and published as an *Index of Economic Articles in Journals* in 34 volumes back to 1886, and distributed electronically as *EconLit* with coverage from 1969. The *Index* and *EconLit* are sold separately from the journals.

This publisher's costs are summarized in figure 1. Some costs seem unlikely to be affected by the digital medium, while others may change significantly. The headquarters function accounts for 27 percent of the AEA's budget. The headquarters maintains the mailing lists, handles the receipts, and does the accounting and legal work. It conducts an annual mail ballot to elect new officers, and organizes an annual meeting that typically draws 8,000 persons.[10,11] The headquarters function seems likely to continue in about its current size as long as the AEA continues as a membership organization, a successful publisher, and a coordinator of an annual meeting.[12] Declining membership or new modes of serving members might lead to reduction in headquarters costs. In the short run, headquarters costs are not closely tied to the number of members or sale of journals.

The AEA's second function is editing, the second block in figure 1. Thirty-six percent of the AEA's annual expenditures goes to the editorial function of its three journals. Eighty-eight percent of the editorial cost is for salaries. The editorial function is essential to maintaining the high production values that are necessary for successful information products.

Operating digitally may provide some cost saving in the editorial function for the *American Economic Review*. The editors could allow manuscripts to be posted on the Internet, referees could access network copies, and dispatch their comments via the network. The flow of some 1,600 referee reports that the AER manages each year might occur faster and at lower cost to both the journals and the referees if the network were used in an effective way.[13] However, the editorial cost will continue to be a significant and essential cost of bringing successful intellectual products to market. Top quality products are likely to have higher editorial costs than lower quality products.

The top two blocks shown in figure 1 describe the 48 percent of the AEA's total budget that goes to printing and mailing. These functions are contracted out, and have recently gone through a competitive bid process. The costs are likely to be near industry lows. The total printing and mailing costs split into two parts. One part doesn't vary with the size of the print run and is labeled as fixed cost. It includes design and typesetting and thus will remain, to a significant degree, as a necessary function in bringing high quality products to market.[14] The variable-cost part of printing and mailing reflects the extra cost of paper, printing, and mailing individual paper issues. These 23 percent of total Association expenditures, $800,000 out of $3.7 million total, might be reduced considerably by using distribution by network. However, as long as some part of the journal is distributed in print, the Association will continue to incur significant fixed costs in printing.

In short, distribution of the journals electronically by network might lower the AEA's expenditures by as much as 23 percent.[15]

Publisher Revenue

Figure 2 summarizes the American Economic Association's revenues in six categories. Thirty-eight percent of revenue comes from individual memberships. Another five percent comes from the sale of advertising that appears in the journals. Nineteen percent comes from the sale of subscriptions, primarily to libraries. Another 19 percent comes from royalties on licenses of the *EconLit* database, most of these royalties come from SilverPlatter, a distributor of electronic databases. Less than half of one percent of revenues come from selling rights to reprint journal articles. Finally, 17 percent of revenues come from other sources, primarily income from the cumulated reserves as well as net earnings from the annual meeting.[16]

Distributing the journals electronically by network seems likely to change the revenue streams. What product pricing and packaging strategies might allow the AEA to sustain the journals? If the journals are to continue to play an important role in the advance of the discipline, then the Association must be assured that revenue streams are sufficient to carry the necessary costs.

If the library subscription includes a license for making the journals available by network to all persons within a campus, then a primary reason for membership in the Association may be lost. With print, the main distinction between the library subscription and the membership subscription is that the member's copy can be kept at hand while the library copy is at a distance and may be in use or lost. With electronic delivery, access may be the same everywhere on the campus network. The license for electronic network distribution may then undercut revenues from memberships, a core 38 percent of AEA revenues.

The demand for advertising in the journals is probably motivated by distribution of journals to individual members. If individual subscriptions lag, then advertising revenue may fall as well. Indeed, one may ask the deeper question of whether ads associated with electronic journals will be salient when the journals are distributed electronically? The potential for advertising may be particularly limited if the electronic journals are distributed through intermediaries. If a database intermediary provides an index to hundreds of journals and provides links to individual articles on demand, advertising revenue may acrue to the database vendor rather than the publisher of the individual journal.

The AEA might see 43 percent of its revenues (the 38 percent from member fees plus the 5 percent from advertising) as vulnerable to being cannibalized by network licensure of its journals. With only a potential 23 percent saving in cost, the Association will be concerned to increase revenues from other sources so as to sustain its journals. The 20 percent shortfall is about $750,000 for the AEA. Here are three strategies: a) charge libraries more for campus-use licenses, b) increase revenues from pay-per-look services, c) enhance services for members so as to sustain member revenues. Each of these strategies may provide new ways of generating revenue from existing readers, but importantly, may attract new readers.

The Campus License

The Association could charge a higher price to libraries for the right to distribute the electronic journals on campus networks. There are about four memberships for each library or other subscription. If membership went to zero because the subscriptions all became campus intranet licenses, then the AEA would need to recoup the revenues from four memberships from each campus license to sustain current revenues. If network distribution lowered AEA costs by 20 percent, then the campus intranet license need only recoup the equivalent of two memberships. Libraries currently pay double the rate of memberships, so the campus intranet license need be only double the current library subscription rate. That is, the current library rate of $140 would need to go to about $280 for a campus-wide intranet license for the three journals.[17] Of course, many campuses have more than one library subscription, say one each in the social science, management, law, and agriculture libraries. The Association might then set a sliding scale of rates from $280 for a small (one library print subscription) campus to $1,400 for a large (five library print subscription) campus.[18] These rates would be the total revenue required by the Association for campus-subscription assuming that the library's print subscriptions are abandoned. A database distributor would add some mark-up.

The campus intranet rate for electronic access is easily differentiated from the print library subscription because it provides a license for anyone on the campus intranet to use the journals in full electronic format. This rate could be established as a price for a new product, allowing the print subscriptions to continue at library rates. Transition from print to electronic distribution could occur gradually with the pace of change set by libraries. Libraries would be free to make separate decisions about adding the campus intranet service and, later, dropping the print subscription.

Individual Association members could continue their print subscriptions as long as they wish, reflecting their own tastes for the print product and the quality of service of the electronic one as delivered. Indeed, individual members might get passwords for direct access to the on-line journals. Some members may not be affiliated with institutions that subscribe to network licenses.

It is possible that the campus intranet license will be purchased by campuses that have not previously subscribed to the AEA's journals. If the institution's cost of participating in network delivery is much less than the cost entailed in sustaining the print subscription, for example, the avoidance of added shelf space as will be discussed below, then more campuses might sign on. This effect may be small for the AEA because it is the premier publisher in economics, but might be significant for other journal publishers.

Pay-Per-Look

The AEA has had minimal revenues from reprints and royalties on copies. Indeed, it pioneered in guaranteeing in each issue of its journals, a limited right to copy for academic purposes without charge.[19] The Association adopted the view that the cost of processing the requests to make copies for class purposes (which it routinely granted without charge), were not worth incurring. By publishing a limited, no-charge right to copy, it saved itself the cost of managing the granting of permissions and saved campuses the cost of seeking them.

With electronic distribution, the campus intranet license will automatically grant permission for the journals to be used in course reserves and in print-on-demand services for classes.

On campuses with too little commitment to instruction in economics to justify a library subscription or a campus intranet license, there may still be occasional interest in use of journal articles. There may be law firms, businesses, consulting enterprises, and public interest groups who occasionally seek information and would value the intensity of exploration found in academic journals. With the ubiquitous Internet, they should be able to search a database on-line for a modest usage fee, identify articles of interest, and then call up such articles in full-image format on a pay-per-look basis. Suppose the Internet reaches a million people who are either on campuses without print library subscriptions today or not on campuses at all, but who would have interest in some occasional use of the academic material. This market represents a new potential source of revenue for the AEA which could be reached by an Internet-based a pay-per-look price.

What rate should the Association set per page to serve the pay-per-look market without unduly cannibalizing the sale of campus intranet licenses? Let's take a one-print library subscription campus rate at $280 per year for access to about 3,500 published pages of journal articles (leaving aside the index and abstracts). One look at each published article page per year at eight cents per page would equal the $280 license. A campus that had a distribution of users that averaged one look at each page would break-even with the campus intranet license with a pay-per-look rate of eight cents per page. This rate is the rate of net revenue to the Association, the database distributor may add a mark-up. For discussion, suppose the database distributor's mark-up is 100 percent. If the Internet users beyond the campus intranet licenses looked at 2 million pages per year at 16 cents per page including fees to the Internet service provider, the Association would recoup nearly a quarter of its lost membership revenue from the intranet licenses from this source.

A critical issue for the emergence of a pay-per-look market is the ability to account for and collect the charges with a low cost per transaction. If accounting and billing costs $10 per hit with hits averaging 20 pages, then the charge might be $14.00 per hit ($10 to the agent, $4 to the AEA). Such a rate compares well with the $30 per exchange of costs incurred in conventional interlibrary loan. Yet such high transactions costs will surely limit the pay-per-look market.

A number of enterprises are offering or plan to offer electronic payment mechanisms on the Internet.[20] In the library world, RLG's WebDOC system may have some of the necessary features. These systems depend on users being registered in advance with the web-bank. As registered users they have accounts and encrypted "keys" that electronically establish their identity to a computer on the net. To make a transaction, a user need only identify herself to the electronic database vendor's computer using the "key" for authentication. The vendor's computer checks the authentication and debits the readers' account at the web-bank. In this fashion, secure transactions may occur over the network without human intervention at costs of a few cents per hit. If such web-banks become a general feature of the Internet, web-money will be used for a variety of purposes. The incremental cost of using them for access to information should be modest and the pay-per-look market gain importance. Mark-ups per transaction might then be quite modest, with gross charges per page in the vicinity of 10 to 20 cents. This rate compares with the four cent per page cost of the *Britannica* when no per page charge is imposed as mentioned in the opening sentence of this essay.

The core idea here is that individual readers make the decisions about when to look at a document under a pay-per-look regime. The reader must face a budget constraint, that is, have a limited set of funds for use in buying information products or other services. The fund might be subsidized by the reader's institution, but the core choices about when to pay and look are made individually. When the core decision is made by the reader with limited funds, then the price elasticity of demand for such services may be high. With a highly elastic demand, even for profit publishers will find that low prices dominate.

Current article fulfillment rates of $10 to $20 could fall by an order of magnitude. The MIT Press offers to deliver individual articles from its electronic journals for $12. EI Village delivers reprints of articles by fax or other electronic means for fees in this range.

Enhanced Member Services

A third strategy for responding to the possible revenue shortfall from the loss of memberships at the AEA would be to enhance membership services. One approach, proposed by Hal Varian, would be to offer superior access to the electronic journals to members only.[21] The electronic database of journal articles might be easily adapted to provide a personal notification to each member as articles of interest are posted. The Association's database service for members might then have individual passwords for members and store profiles of member interests so as to send e-mail notices of appropriate new postings. The members' database might also contain ancillary materials, appendices to the published articles with detailed derivations of mathematical results offered in software code (for example, as *Mathematica* notebooks), copies of the numerical data sets used in empirical estimation, or extended bibliographies. The members' database might support monitored discussions of the published essays, allowing members to post questions and comments and an opportunity for authors to respond if they wish. These enhancements generally take advantage of the personal relationship a member may want to have with the published literature, a service not necessarily practical or appropriate for libraries.

Indeed, one divide in the effort to distinguish member from library access to the journal database is whether the enhancement would have value to libraries if offered. Libraries will be asked to pay a premium price for a campus intranet license. They serve many students and faculty who are not currently members of the AEA and who are unlikely to become members in any event; for example, faculty from disciplines other than economics. Deliberately crippling the library version of the electronic journals by offering lower resolution pages, limited searching strategies, a delay in access, or only a subset of the content, will be undesirable for libraries and inconsistent with the Association's goal of promoting discussion of economics. However, there may me some demand for lower quality access at reduced prices. The important point is that for membership to be sustained, it must carry worthwhile value when compared to the service provided by the campus license.

Another approach is simply to develop new products that will have a higher appeal to members than to libraries. Such products could be included in the membership fee, but offered to libraries at an added extra cost. One such product would be systematic access to working papers in economics. Indices, abstracts, and in some cases, the full-text of working papers are available without charge at some sites on the World Wide Web today. The Association might ally itself with one of these sites, give the service an official status, and invest in the features of the working paper service to make it more robust and useful. Although freebie working paper services are useful, an enhanced working paper service for a fee (or as part of membership) might be much better.[22]

To the extent that enhanced services can sustain memberships in the face of readily available campus intranet access to journals, the premium for campus intranet access could be lower.

The AEA might offer a discount membership rate to those who opt to use the on-line version of the journals in lieu of receiving print copies. Such a discounted rate would reflect not only the Association's cost saving with reduced print distribution but also the diminished value of membership given the increased prospect of campus intranet licenses.

To the extent that the pay-per-look market generates new revenue, then the campus intranet rate could also be less. The total of the Association's revenues need only cover its fixed and variable costs. (The variable cost may approach zero with electronic distribution.) If membership revenues dropped by two-thirds and pay-per-look generated one-quarter of the gap, then the premium rate for the campus intranet license need be only one-third to one-half above current rates, say, $200 for a one-print subscription campus to $1,000 for a five-print library subscription campus (net revenue to the Association after the net distributor's mark-up).

Other Publishers

At the other end of the publishing spectrum from the AEA are those producing low volume publications. Some titles have few personal subscriptions and depend primarily on library subscriptions that are already at premium rates. For these titles, replacing the print subscription with an intranet license will simply lower costs. The Johns Hopkins University Press offers its journals electronically at a discount in substitution for the print.

Some titles may have mostly personal subscriptions with no library rate, including popular magazines like the *Economist*. Such publications might simply be offered as personal subscriptions on the Internet with an individual password for each subscriber. The distribution by network would lower distribution costs and so ought to cause the profit maximizing publisher to offer network access to individuals at a discount from the print subscription rate. Such a publication may not be available by campus intranet license.

The *Journal of Statistics Education* (JSE) is distributed via the Internet without charge. It began with an NSF/FIPSE grant to the North Carolina State University in 1993. The JSE receives about 40 manuscripts per year and, after a peer review, publishes about 20 of them.[23] The published essays are posted on a web site and a table of contents and brief summaries are dispatched by e-mail to a list of about 2,000 interested persons. JSE's costs amount to about $25,000 per year to sustain the clerical work necessary to receive manuscripts, dispatch them to suitable referees, receive referee reports, and return them to the author with the editor's judgment. The JSE also requires a part-time system support person to maintain the server that houses the journal. The JSE has not charged for subscriptions, receives no continuing revenue, and needs about $50,000 per year to survive. Merger with a publisher of other statistics journals may make sense, allowing the JSE to be bundled in a larger member service package. Alternatively, it might begin to charge a subscription fee for individuals and a campus license rate for libraries. Making the transformation from a no-fee to a fee-based publication may prove difficult. A critical issue is how much fixed cost is necessary to maintain reasonable production values in a low volume publication. At present, JSE is seeking a continuing source of finance.

In general, a publisher will consider three potential markets: (1) the campus intranet license/library sale, (2) the individual subscription, and (3) the pay-per-look/individual article sale. These three markets might be served by one title with shared fixed costs. The issue of whether to offer the title in each market and at what price will reflect the incremental cost of making the title available in that market, the elasticity of demand in each market, and the cross price elasticities between markets. For example, the price of the campus license will have an effect on individual subscription sales, and the price of the individual subscriptions will have an effect on the sale of individual articles, and vice versa. The more elastic the demands, the lower the prices, even for for-profit publishers. With higher substitution between the three forms, the closer the prices will be across the three forms.[24]

Economies of Scope

To this point, the analysis applies essentially to one journal at a time, as though the journal were the only size package that counted. In fact, of course, the choice of size of package for information could change. Two centuries ago, the book was the package of choice. Authors generally wrote books. Libraries bought books. Readers read books. In the last fifty years, the size of package shifted to the journal in most disciplines. Authors write smaller packages, that is, articles, and get their work to market more quickly in journals. The elemental information product has become more granular. Libraries commit to journals and so receive information faster and at lower cost per unit. In deciding what to read, readers depend on the editors' judgment in publishing articles. In short, libraries buy bigger packages, the journals, while authors and readers work with smaller units, the articles.

With electronic distribution, the library will prefer to buy a still larger package, a database of many journals. A single, large transaction is much less expensive for a library to handle than the multiple, small transactions. Managing many journal titles individually is expensive. Similarly, readers may prefer access to packages smaller than journal articles. They are often satisfied with abstracts. The electronic encyclopedia is attractive because it allows one to zip directly to a short, focused package of information with links to more. Authors, then, will be drawn to package their products in small bundles embedded in a large database with links to other elements of the database with related information. Information will become still more granular.

If the database becomes the dominant unit of trade in academic information, then those with better databases may thrive. The JSTOR enterprise appears to have recognized the economies of scope in building a database with a large quantity of related journal titles. JSTOR is a venture spawned by the Mellon Foundation to store archival copies of the full historic backfiles of journals and make them available by network. The core motive is to save libraries the cost of storing old journals. JSTOR plans to offer 100 journal titles within a few years. Some of the professional societies, for example, psychology and chemistry, exploit economies of scope in the print arena by offering dozens of journal titles in their disciplines. Elsevier's dominance in a number of fields is based in part on the exploitation of scope with many titles in related subdisciplines. The emergence of economies of scope in the electronic arena is illustrated by Academic Press's offer to libraries in Ohio Link. For ten percent more than the cost of the print subscriptions the library had held, it could buy electronic access to the full suite of Academic Press journals electronically on Ohio Link.

To exploit the economies of scope, the electronic journal might begin to include hot links to other materials in the database. The electronic product would then deliver more than the print version. Links to other web-sites is one of the attractive features of the web-version of the *Encyclopedia Britannica*. An academic journal database could invite authors to include the electronic addresses of references and links to ancillary files. Higher quality databases will have more such links.

The American Economic Association eschews scope in the print arena, preferring instead to let a hundred flowers bloom and to rely on competition to limit prices. Its collection of three journals does not constitute a critical mass of journal articles for an economics database and so it must depend on integration with other economics journals at the database level. The Johns Hopkins University Press's Muse enterprise suffers similar lack of scope. Although it has 45 journal titles, they are scattered among many disciplines and do not, collectively, reach critical mass in any field.

The emergence of more powerful, network-based working paper services seems likely to lower the cost of the editorial process, as mentioned above. A common, well-managed electronic working-paper service might make the cost of adding a journal title much lower than starting a title from scratch without access to electronic working papers. The enterprise that controls a capable working paper service may well control a significant part of the discipline and reap many of the advantages of scope in academic publishing.

In fact, a capable electronic working paper service could support multiple editors of a common literature. One editor might encourage an author to develop a work for a very sophisticated audience and publish the resulting work in a top academic journal. Another editor might invite the author to develop the same ideas in a less technical form for a wider audience. Both essays might appear in a common database of articles and link to longer versions of the work, to numerical data sets, bibliographies, and other related material. The published essays will then be front-ends to a deeper literature available on the Net.

Rents

In addition to limiting the number of journals it produces, the American Economic Association differs from many publishers by emphasizing low cost. The price of its journals is less than half the industry average for economics journals, and the differential between library and individual rates is low.[25] If the AEA's goal were to maximize profit, it could charge authors more, charge members and libraries more, make more revenue from its meetings, and launch more products to take advantage of its reputation by extending its scope. The rents available in this marketplace are then left to the authors, members, libraries, and competing publishers. The AEA is not maximizing its institutional rents.

Other non-profit publishers may seek higher revenues, to capture more of the available rents, and use the proceeds to generate more products and association services. Lobbying activities, professional certification and accreditation, more meetings, and more journals are common among professional societies.

Many for-profit publishers seek to maximize the rents they can extract from the marketplace for the benefit of their shareholders. In considering how to package and price electronic products, the for-profit publishers will continue to be concerned with finding and exploiting the available rents. The profit maximizing price for a journal is determined by the price elasticity of demand for the title and the marginal cost of producing it. With convenient network access, there may be an increase in demand that would allow a higher price, other things equal. How the price elasticity of demand might change with network access is unknown. The fall in marginal cost with electronic distribution need not lead to a lower price.

One might then ask how a shift to electronic publishing may affect the size of the rents and their distribution. A shift to the database as the optimal size package with falling marginal costs would seem both to increase the size of potential rents and to make easier their exploitation for profit. Suppose control of a powerful working paper service gives a significant cost advantage to journal publishers. Suppose further that academic institutions find major advantages in subscribing to large databases of information rather than making decisions about individual journal titles. The enterprise that controls the working paper service and the database of journals may then have considerable rent capturing ability. The price elasticities of demand for such large packages may be low and the substitutes poor, and so the mark-ups over costs may be substantial. The possibility of a significant pay-per-look market with high price elasticity of demand might cause the profit maximizing price to be lower. The possibility of self-publication at personal or small scale web sites offers a poor substitute to integration in a database because web search engines are unlike to point to them appropriately.

Library

In contemplating how to take advantage of electronic publications, universities and their libraries face two problems. First, they face decisions about scaling back costly conventional operations so as to make resources available for acquiring electronic licenses. Second, the cost savings occur in a variety of ways, each with its own history, culture, and revenue sources. Although many boards of trustees and their presidents might like all of the funds within their institutions to be fungible, in fact they face limitations on their ability to reduce expenditures in one area so as to spend more in another. If donors or legislatures are more willing to provide funds for buildings than for electronic subscriptions, then the dollar cost of a building may not be strictly comparable to the dollar cost of electronic subscriptions. Universities are investing more in campus networks and computer systems and are pruning elsewhere as the campuses become more digital. The following paragraphs consider how conventional operations might be pruned so as to allow more expenditure on electronic information products.

Conventional Library Costs

It is possible that some universities will view electronic access to quality academic journals as sufficiently attractive to justify increasing their library budget to accommodate the electronic subscriptions when publishers seek premium prices for electronic access. Some universities place particular emphasis on being electronic pioneers and seem willing to commit surprising amounts of resources to such activities. Other universities owe a debt to these pathfinders for sorting out what works. However, for most institutions, the value of the electronic journals will be tested by middle management's willingness to prune other activities so as to acquire more electronic journals. The library director is at the front line for such choices and an understanding of the basic structure of the library's expenditures will help define the library director's choices.

Figure 3 provides a summary picture of the pattern of costs in conventional academic libraries. The top four blocks correspond to the operating budgets of the libraries. Acquisitions account for about a third of the operating budget. To give a complete picture, the bottom section of the figure also accounts for the costs of library buildings. The cost of space is treated as the annual lease value of the space including utilities and janitorial services. The total of the operating budget plus the annualized cost of the building space represents a measure of the total institutional financial commitment to the library.

Library management typically has control only of the operating budget. Let's suppose that, on average, campus intranet licenses to electronic journals come at a premium price, reflecting both the electronic database distributor's costs as well as adjustments in publishers pricing behavior as discussed above. The library, then, confronts a desire to increase its acquisition expenditure, possibly as much as doubling it.

A first choice is to prune expenditures on print so as to commit resources to digital materials. Some publishers offer lower prices for swapping digital for paper and in this case, swapping improves the libraries budget. Some publishers may simply offer to swap digital for print at no change in price. However, many may expect a premium gross price for digital access on the campus intranet. The library manager may seek to trim other acquisition expenditures so as to commit to more digital access. For several decades, academic libraries have been reducing the quantity of materials acquired so as to adjust to increases in prices. The possibility of substantial cuts in the quantity of acquisitions so as to afford a smaller suite of products in electronic access seems unappealing and so may have limited effect.

A second possible budget adjustment is to prune technical service costs. The costs of processing arise from the necessity of tracking the arrival of each issue, claiming those that are overdue, making payments, adjusting catalog records, and periodically binding the volumes. If the electronic journal comes embedded in a database of many journals, the library can make one acquisition decision and one payment. It need have little concern for check-in and the claiming of issues. Testing the reliability of the database will be a concern but presumably large database providers have a substantial incentive to build in considerable redundancy and reliability and will carefully track and claim individual issues, once for all. The library will avoid binding costs. The library will likely have some interest in building references to the electronic database into its catalog. Perhaps the database vendor will provide suitable machine readable records to automate this process.

A third possibility is the library's public service operations. Until a substantial quantity of materials are available and widely used via network, the demand for conventional library hours, reference, and circulation services may change only modestly. In 1996, a third to a half of the references in my students' essays were to World Wide Web sources. However, these sources generally complemented conventional sources rather than being substitutes for them. As front-line journals become commonly accessible by campus networks, the demand for conventional library services may decline. For example, campuses that operate departmental and small branch libraries primarily to provide convenient access to current journals for faculty might be more likely to consolidate such facilities into a master library when a significant number of the relevant journals are available on the Net. These changes are likely to take a number of years to evolve.

A fourth possibility concerns the cost of library buildings. When journals are used digitally by network, the need for added library space declines. Libraries will need less stack space to hold the addition of current volumes. In many larger libraries, lesser used, older volumes are currently held in less expensive, off-site facilities, with new volumes going into the prime space. The marginal stack space, then, is off-site, with costs of perhaps $0.30 per volume per year as a continuing cost for sustaining the perpetual storage of the added volumes.[26] Replacing a 100 year run of a journal with an electronic backfile ought to save about $30 per year in continuing storage costs at a low-cost, remote storage facility. Reductions in the extent of processing and in public services will also reduce requirements for space.

The library building expenses typically do not appear in operating budgets, so saving space has no direct effect on the library budget. The capital costs of buildings are frequently raised philanthropically or paid through a state capital budget, keeping the costs out of the university current accounts. Even utilities and janitorial services may appear in a general university operating budget rather than appearing within the library account. Savings in building costs will accrue to those who fund capital projects and to university general budgets, but often, not to the library operating budget. University presidents and boards may redirect their institutions' capital funds to more productive uses. Of course, the interests of philanthropy and the enthusiasm of state legislators may pose some limit on the ability to make such reallocations. Moreover, library building projects occur relatively infrequently, say every 25 years or so. The savings in capital may not be apparent for some time, or indeed, ever if capital budgets are considered independently of operating budgets. Library buildings, particularly the big ones in the middle of campuses, come to play a symbolic role, an expression of the university's importance, a place of interdisciplinary interaction, a grand presence. Because symbols are important, the master library facility will continue to be important. The marginal savings in building expense will probably be in compact or remote storage facilities and in departmental and smaller branch libraries. Digital access ought then to save the larger campus community some future commitment of capital, but the savings will be visible mostly to the president and board.

A fifth possibility is savings in faculty subscriptions. In law, business, and other schools where faculty have university expense accounts, faculty may be accustomed to paying for personal subscriptions to core journals from the accounts. If the university acquires a campus-wide network license for such journals, the faculty members may rely on the campus license and deploy their expense accounts for other purposes. By adjusting the expense account downward in light of the offering of campus licenses for journals, the university may reclaim some of the cost of the journals. On those campuses and in those departments where faculty members do not have expense accounts and where personal copies of core journals are necessary for scholarly success, the faculty salaries might be adjusted downward over a course of time to reflect the fact that faculty may use the campus license rather than pay for personal subscriptions. Indeed, when the personal subscriptions are not deductible under federal and state income taxes, the cost of subscriptions to the faculty in after tax dollars may be greater than the cost to the university using before tax dollars. As a result a shift to university site licenses for core journals should be financially advantageous for faculty and the university.

In sum, the university may find a number of ways to economize by shifting to digital journals distributed by network. Although direct subscription prices may go up in some cases, the university may trim technical and public services, save space, and offer more perquisites to faculty at some saving in cost.

Electronic Agent

Publishers could establish their own digital distribution function by creating a Universal Resource Locator (URL) for each title. The publisher would deal directly with libraries and individual readers. For a number of reasons, the publisher is likely to prefer to work with an agent for electronic distribution. Just as the typesetting and printing is usually performed by contractors, so the design and distribution of electronic products is likely to involve specialized agents. However, the role of electronic distribution agent is becoming more important than that of the printer for two important reasons. The first arises because of economies of scale in managing access to electronic services. The second concerns the potential advantages of integrating individual journals into a wider database of academic information. The electronic agent accepts materials, say journal titles, from publishers and mounts them on electronic services to be accessed by the Internet. The agent captures economies of scale in maintaining the service, in supporting a common payment mechanism, a common search interface and search engine, and may take other steps to integrate articles and journal titles so that the whole is greater than the sum of the parts.

OCLC was an early entrant in the market for electronic distribution of academic journals with *Online Clinical Trials*. *Online Clinical Trials* was priced at $220 for institutions and $120 for individuals.[27] OCLC is shifting to a World Wide Web interface in January, 1997 and hopes to offer more than 250 journal titles soon. OCLC's new approach offers publishers the opportunity to sell electronic access to journals by both subscription and pay-per-look.[28] It charges libraries an access fee based on the number of simultaneous users to be supported and the number of electronic journals to which the library subscribes. Libraries buy subscriptions from publishers. Publishers may package multiple titles together and set whatever rates they choose. The following discussion puts the strategies of OCLC and other electronic agents in a broader context.

Storage and Networks

With electronic documents, there is a basic logistical choice. A storage intensive strategy involves using local storage everywhere. In this case, the network need not be used to read the journal. At the other extreme, the document might be stored once-for-the-world at a single site with network access used each time a journal is read. Between these two extremes, there is a range of choices. With the cost saving of fewer storage sites comes the extra cost of increased reliance on data communication networks.

Data storage is an important cost. Although the unit costs of digital storage have fallen and will continue to fall sharply through time, there is still a considerable advantage to using less storage. Data storage systems involve not simply the storage medium itself, but a range of services to keep the data on-line. A data center typically involves sophisticated personnel, back-up and archiving activities, and the cost of upgrading software and hardware. If ten campuses share a data storage facility, the storage cost per campus should be much less than if each provides its own. Having one storage site for the world might be the lowest storage cost per campus overall.

To use a remote storage facility involves data communication. The more remote the storage, the greater the reliance on data networks. A central problem for data communication is congestion. Data networks typically do not involve traffic-based fees. Indeed, the cost of monitoring traffic so as to impose fees may be cost prohibitive. Monitoring network traffic so as to bill to individuals on the basis of use would require keeping track of the origin of each packet of data and accounting for it by tallying a register that notes source, time, and date. Because even simple mail messages may be broken into numerous packets for network shipment, the quantity of items to be tracked is much more numerous than tracking telephone calls. If every packet must go through the toll plaza, the opportunity for delay and single points of failure may be substantial. Because each packet may follow a different route, tracking backbone use with a tally on each leg would multiply the complexity. Traffic-based fees seem to be impractical for the Internet. Without traffic-based fees, individual users do not face the cost of their access. Just as with urban highways at rush hour, each individual sees only his or her own trip, not the adverse effect of his or her trip in slowing others down. An engineering response to highway congestion is often to build more highways. Yet, the added highways are often congested as well. In data networking, an engineering solution is to invent a faster network. Yet, individuals deciding to use the network will see only their personal costs, and so have little incentive to economize. The demand for bandwidth on networks will surely grow with the pace of faster networks, for example, with personal videophones and other video intensive applications. Without traffic-based pricing, congestion will be endemic in data networks.

Another response to network congestion is to build private networks with controlled access. Building networks dedicated to specific functions seems relatively expensive, but may be necessary to maintain a sufficient level of performance. Campus networks are private, and so access can be controlled. Perhaps investments in networking and technical change can proceed fast enough on individual campuses to allow the campus network to be reliable enough for access to journals and other academic information.

As the telephone companies have launched data network services, they seem likely to introduce time-of-day pricing. Higher rates in prime time and higher rates for faster access speeds are first steps in giving incentives to economize the use of the network and so to reduce congestion. America On Line (AOL) ran into serious difficulty when in late 1996 it shifted from a per hour pricing strategy to a flat monthly rate to match other Internet service providers. AOL was swamped with peak period demand, demand it could not easily manage. The long distance telephone services seem to be moving to simpler pricing regimes, dime-a-minute, for example. The possibility of peak period congestion, however, likely means that some use of peak period pricing in telephones and in network services will remain desirable. In the end, higher education's ability to economize on data storage will depend on the success of the networks in limiting congestion.

Some milestones in the choice of storage and networks are illustrated along the horizontal margin of figure 4. The rapid growth of the World Wide Web in the last couple of years has represented a shift toward the right along this margin, with fewer storage sites and more dependence on data communication. The World Wide Web allows a common interface to serve many computer platforms, replacing proprietary tools. Adobe's Portable Document Format (PDF) seems to offer an effective vehicle to present documents in original printed format with equations, tables, and graphics, yet allow text searching and hypertext links to other websites. The software for reading PDF documents is available without charge, compatible with many web browsers, and allows local printing. Some of the inconveniences of older network-based tools are disappearing.

The electronic agent may have an advantage over either the publisher or the library in taking advantage of the rightward shift. That is, the electronic agent may acquire rights from publishers and sell access to libraries, while taking responsibility for an optimal choice of storage sites and network access. Storage might end up in a low cost location with the electronic agent responsible for archiving the material and migrating the digital files to future hardware and software environments.

Integration into a Database

The second advantage for an electronic agent is in integrating individual journal titles and other electronic materials into a coherent database. The vertical margin of figure 4 sketches a range of possibilities. At root, a journal title stands as a relatively isolated vehicle for the distribution of information. In the digital world, each title could be distributed on its own CD or have its own

BEST COPY AVAILABLE

Universal Resource Locator on the web. Third party index publishers would index the contents and provide pointers to the title and issue, and perhaps to the URL. Indeed, the pointer might go directly to an individual article.

However, relatively few scholars depend on a single journal title for their work. Indeed, looking at the citations shown in a sampling of articles of a given journal reveals that scholars typically use a range of sources. A database that provides coherent access to several related journals, as in the second tier of figure 4, offers a service that is more than the sum of its parts.

At yet a higher level, an agent might offer a significant core of the literature in a discipline. The core of journals and other materials might allow searching by words and phrases across the full content of the database. The database then offers new ways of establishing linkages.

At a fourth level, the organizing engine for the database might be the standard index to the literature of the discipline, such as *EconLit* in economics. A search of the database might achieve a degree of comprehensiveness for the published literature. A significant fraction of the published essays might be delivered on demand by hitting a "fulfill" button. Fulfillment might mean delivery of an electronic image file via network within a few seconds or delivery of a facsimile within a few minutes or hours.

At a fifth level, the database might include hot-links from citations in one essay to other elements of the database. The database might include the published works from journals with links to ancillary materials, numeric data-sets, computer algorithms, an author's appendices discussing methods and other matters. The database might invite commentary and so formal publications might link to suitably moderated on-line discussions.

The electronic agent may have an advantage over publishers who offer only individual journal titles in integrating materials from a variety of sources into a coherent database. The agent might set standards for inclusion of material that specifies metatags and formats. The agent might manage the index function, indeed, the index might be a basis for forward integration with database distribution as Engineering Information has done. This issue is discussed more fully below.

Integration of diverse materials into a database is likely to come with remote storage and use of networks for access. Integrating the material into a database by achieving higher levels of coherence and interaction among diverse parts may be at lower cost for an electronic agent than for publishers of individual journals or for individual libraries. The agent is able to incur the cost of integration and storage once for the world.

Agent's Strategy

Given the interest of publishers in licensing their products for campus intranets and the universities' interest in securing such licenses, there is opportunity for enterprises to act as brokers, to package the electronic versions of the journals in databases and make them accessible, under suitable licenses, to campus intranets. The brokers may add a mark-up to reflect their cost of mounting the database. The size of the mark-up will reflect the extent of integration as well as the choice of storage strategy.

SilverPlatter became the most successful vendor of electronic index databases, making them available on compact disks for use on campus intranets with proprietary software. OCLC plays an important role in offering such databases from its master center in Ohio. A number of other vendors have also participated in the index market and are likely to seek to be brokers for the electronic distribution of journals. Ovid is a third vendor, one that supports sophisticated indexing that integrated full-text with standard generalized mark-up language (SGML) and hypertext mark-up languge (HTML) tagging.

A core strategy will probably be to mount the database of journals on one or more servers on the World Wide Web, with access limited to persons authorized for use from licensed campuses or through other fee-paid arrangements. This strategy has three important parts, the database server, the Internet communication system, and the campus network.

The advantage of the World Wide Web approach is that the data can be made accessible to many campuses with no server support on any campus. A campus intranet license can be served remotely, saving the university the expense of software, hardware, and system support for the service.

The risk of the Web strategy is with the Internet itself and its inherent congestion. OCLC used a private data communication network so as to achieve a higher level of reliability than the Internet and will do the same to assure high quality TCP/IP (the Internet Protocol) access. Some campuses may prefer to mount database files locally, using CD-ROMs and disk servers on the campus network. Some high intensity campuses may prefer to continue to mount the most used parts of databases locally even at extra cost, as a method of ensuring against deficiencies in Internet services.

The third element after storage and the Internet is the campus network. Campus networks continue to evolve. Among the hundred universities seeking to be top-ten universities, early investment in sophisticated networking may play a strategic role in the quest for rank. On such campuses, network distribution of journals should be well supported and popular. Other campuses will follow with some lag, particularly where funding depends primarily on the public sector. Adoption within ten years might be expected.[29]

The electronic agent, then, must choose a strategy with two elements, a storage and network choice and an approach to database integration.

Journal publishers generally start at the bottom left, the closest to print. They could make a CD and offer it as an alternative to print for current subscribers. The AEA offers the *Journal of Economic Literature* on CD instead of print for the same price.

Moves to the upper left seem to be economically infeasible. Integrating more materials together increases local storage costs and so tilts the storage-network balance toward less storage and more network. With more data integration, the agent's strategy will shift to the right.

Moves to the lower right with reduced storage costs and more dependence on networks should involve considerable cost savings but run risks. One risk is of network congestion. A second is of loss of revenues because traditional subscribers drop purchases in favor of shared network access. The viability of these strategies depends on the level of fees that may be earned from network licenses or pay-per-look.

Moves along the diagonal up and to the right involve greater database integration with cost savings from lower storage costs and more dependence on networks. The advantage of moves upward and to the right is the possibility that integration creates services of significantly more value than replicating print journals on the Internet. When database integration creates significantly more value, subscribers will be willing to pay premium prices for using products with remote storage with networks. Of course, network congestion will remain a concern.

A move toward more database integration raises a number of interesting questions. The answers to these questions will determine the size of the mark-up by the electronic agent. How much should information from a variety of sources be integrated into a database with common structure, tags, and linkages? For a large database, more effort at integration and coherence may be more valuable. Just how much effort, particularly how much hand effort, remains an open question. If the electronic agent passively accepts publications from publishers, the level of integration of materials may be relatively low. The publisher may provide an abstract and metatags and might provide Universal Resource Locators for linking to other network sites. The higher level of integration associated with controlled vocabulary indexing, and a more systematic structure for the database than comes from journal titles would seem to require either a higher level of handwork by an indexer or the imposition of standard protocols for defining data elements. Is a higher level of integration of journal material from a variety of sources sufficiently valuable to justify its cost? The index function might be centralized with storage of individual journals distributed around the net. Physical integration of the database is not necessary to logical integration, but will common ownership be necessary to achieve the control and commonality necessary for high levels of integration?

A second question concerns how an agent might generate a net revenue stream from its initial electronic offerings sufficient to allow it to grow. The new regime will not be borne as a whole entity, rather it will evolve in relatively small steps. Each step must generate a surplus to be used to finance the next step. Early steps that generate larger surpluses seem likely to define paths that are more likely to be followed. Experimentation with products and prices is already underway. Those agents finding early financial success are likely attract publishers and libraries, and to be imitated by competitors.

JSTOR has captured the full historic run of a significant number of journals, making the promise of 100 titles in suites from major disciplines within three years. However, it does not yet have a program for access to current journals. Its program then is primarily to replace archival storage of materials libraries may or may not have already acquired in print.

OCLC's approach is to sell libraries access services while publishers sell subscriptions to the information. The publisher can avoid the cost of the distribution in print, a saving if the electronic subscriptions generate sufficient revenue. The unbundling of access from subscription sales allows the access to be priced on the basis of simultaneous users, that is akin to the rate of use, while the information is priced on the basis of quantity and quality of material made available. Of course, the information may also be priced on a pay-per-look basis and so earn revenue as it is used. What mix of pay-per-look and subscription sales will ultimately prevail is an open question.

A third question is whether publishers will establish exclusive arrangements with electronic agents, or whether they will offer non-exclusive licenses so as to sustain competition among agents. Some publishers may prefer to be their own electronic agents, retaining control of the distribution channels. If database integration is important, this strategy may be economic only for relatively large publishers with suites of journals in given disciplines. Many publishers may choose to distribute their products through multiple channels both to capture the advantages of more integration with other sources, but also to promote innovation and cost savings among competing distributors.

As the electronic agents gain experience and build their title lists, competition among them should drive down the mark-ups for electronic access. If the store-once and network strategy bears fruit, the cost saving in access should be apparent. If higher levels of database integration prove to be important, the cost savings may be modest. Cost savings here are in terms of units of access. As the cost of access falls, the quantity of information products used may increase. The effect on total expenditure, the product of unit cost and number of units used, is hard to predict. If the demand for information proves to be price elastic, then as unit costs and unit prices fall, expenditures on information will increase.

The electronic agents will gather academic journals from publishers and distribute them in electronic formats to libraries and others. They will offer all available advantages of scale in managing electronic storage, optimize the use of networks for distribution, offer superior search interfaces and engines, and take steps to integrate materials from disparate sources into a coherent whole. The agent will be able to offer campus intranet licenses, personal subscriptions, and pay-per-look access from a common source. The agent may manage sales, accounting, billing, and technical support. Today, agents are experimenting with both technical and pricing strategies. It remains to be seen whether single agents will dominate given content areas, whether major publishers can remain apart, or whether publishers and universities can or should sustain a competitive market among agents.

Conclusion

Higher education faces a significant challenge in discovering what academic information will succeed on the Net. In 1996, the MIT Press launched *Studies in Nonlinear Dynamics and Econometrics* (SNDE), one of six titles that the Press distributes by network. The price per year is $40 for individuals and $130 for libraries.[30] MIT's strategy seems to be to launch titles in disciplines where an electronic journal has some extra value, for example, including links to computer code and data sets.[31] The rates for the journals seem to be well below those quoted by OCLC's electronic journal program and lower than at least some new print journals. The cost of launching a new journal electronically seems to be falling. It remains to be seen whether the electronic journals will attract successful editors and valued manuscripts from authors, but the venture shows promise. The number and quality of electronic journals continues to grow. MIT has decided to forgo the use of an electronic agent and so depend only on conventional, independent indexing services for database integration, an incremental approach. Yet, the potential seems greater than an individual journal title reveals.

When Henry Ford launched the first mass produced automobile, he chose a design that carried double the load, went three times farther, and four times faster than the one-horse buggy it replaced, and yet was modestly priced. Successful digital information products for academia seem likely to exploit the inherent advantages of the digital arena, the timeliness, the sophisticated integration of new essays into the existing stock, the links from brief front-end items to more elaborate treatment, the opportunity to interact with the material by asking for "fulfillment," "discussion," and the "underlying data." Network delivery will make possible both the campus intranet license and the sale of information on a pay-per-look basis. It will allow the material to be more readily consulted in circles beyond the academy.

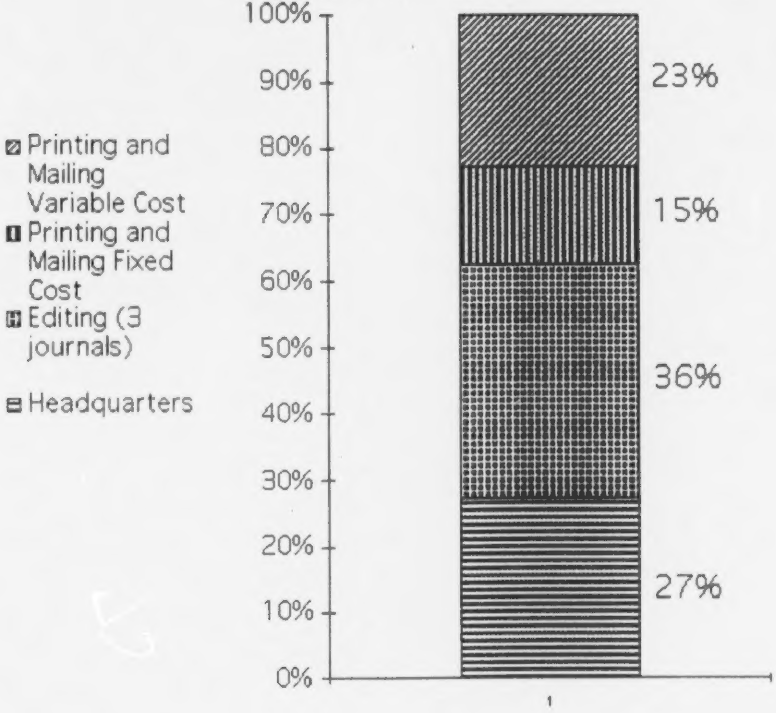
Electronic agents will play significant new roles as intermediaries between publishers and campuses by handling the electronic storage and distribution, and by integrating material into a more coherent whole. Universities and their libraries will make adjustments in operations so as to expend less on conventional activities and more on digital communication.

Of course, there are unknowns. Agents and publishers will experiment to discover optimal pricing strategies. Agents will explore different ways of storing and delivering electronic products and different approaches to integration. Campuses and libraries will consider just what extra dimensions of service are worth their price. The process here is one of bringing order, meaning, and reliability to the emerging world of the Internet, of discovering what sells and what doesn't.

In the end, universities should be drawn to the electronic information services because of their superiority in instruction, their reach beyond the academy, and their power in the creation of new ideas. American higher education is largely shaped by competitive forces, the competition for faculty, students, research funding, public, and philanthropic support. In different ways, the private and public sector, the large institutions and the small, the two-year and four-year institutions share the goal of doing a better, more cost effective job of expanding the human potential. When artfully done, the digital sharing of ideas seems likely to expand that potential significantly.

Figure 1

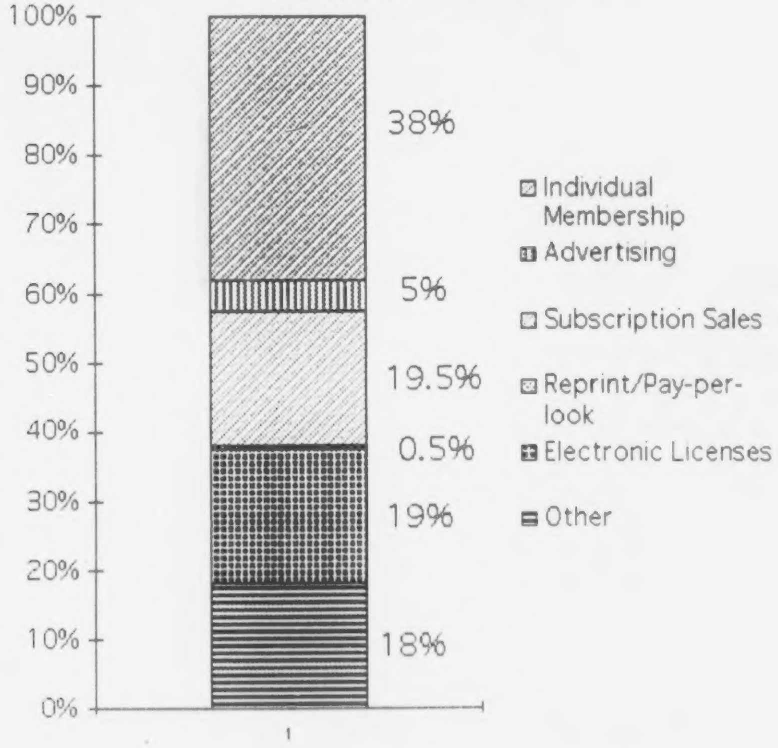
American Economic Association Expenses 1995



Source: Elton Hinshaw, "Treasurer's Report," *American Economic Review*, May, 1996 and unpublished reports.

Return to the Document

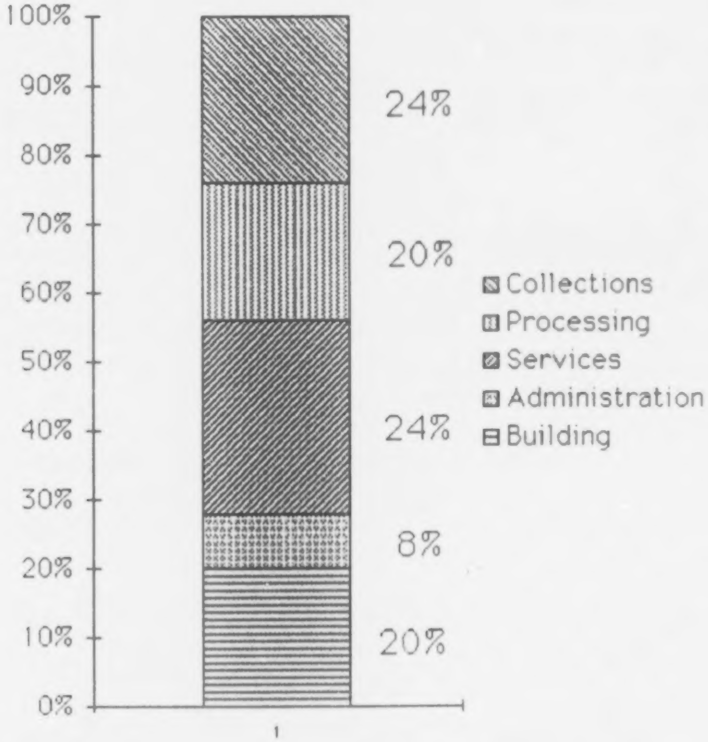
Figure 2
American Economic Association
Revenues 1995



Source: Elton Hinshaw, "Treasurer's Report," *American Economic Review*, May, 1996 and unpublished reports.

Return to the Document

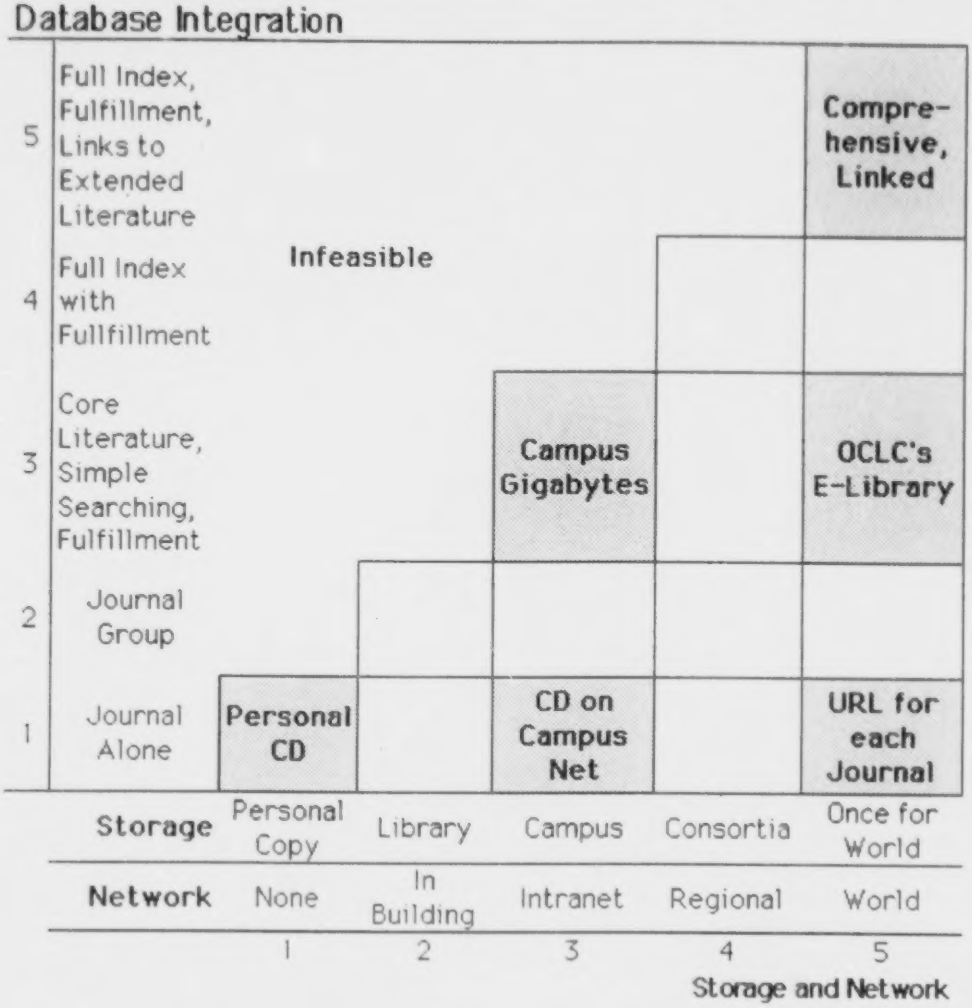
Figure 3
Conventional Library Costs



source: Heuristic characterization based on Association of Research Libraries Annual Statistical Survey on expenditures on materials and operating budgets, and the author's own studies of library space and technical service costs

Return to the Document

Figure 4
Network Intensity and Database Integration


Database Integration
5
Full Index, Fulfillment, Links to Extended Literature
4
Full Index with Fullfillment
3
Core Literature, Simple Searching, Fulfillment
2
Journal Group
1
Journal Alone
Infeasible
Compre-hensive, Linked
Campus Gigabytes
OCLC's E-Library
Personal CD
CD on Campus Net
URL for each Journal
Storage
Personal Copy
Library
Campus
Consortia
Once for World
Network
None
In Building
Intranet
Regional
World
1
2
3
4
5
Storage and Network

Return to the Document

FOOTNOTES:

---

[1] I appreciate the help of Elton Hinshaw and the American Economic Association in understanding its operations and the comments of Paul Gherman, David Lucking-Reiley, and Flo Wilson on an earlier draft of this essay.

[2] Shirley Baker, talk at Washington University, November, 1996.

[3] Robin Frost, "The Electronic Gutenberg Fails to Win Mass Appeal," *Wall Street Journal*, November 21, 1996, p. B6. Project Gutenberg was a twenty-five year effort led by Michael S. Hart at the University of Illinois to create, store, and make accessible ASCII files of public domain materials from the Constitution, the Bible, Shakespeare, and beyond.

[4] Stephen Burd, "President Pushes Tax Breaks to Help Families Afford College," *Chronicle of Higher Education*, January 17, 1997, p. A33.

[5] www.ei.org

[6] http://xxx.lanl.gov/

[7] http://scriptorium.lib.duke.edu/papyrus/ offers 1,373 images of Egyptian papyri with a significant database of descriptive textual material.

[8] http://fairmodel.econ.yale.edu/

[9] http://gdbwww.gdb.org/

[10] The headquarters publishes *Job Openings in Economics* (JOE) seven times a year with nearly 1,500 job announcements. In 1995, JOE had about 4,000 subscribers and generated about $41,000 of revenue with a base rate of $15 per year ($7.50 for students, $25 for nono-members and institutions). The sum of monthly printing and mailing cost was associated with the number of copies produced and the number of pages per copy for 1995 and 1996 as follows (with t-ratios in parenthesis):

```
Print & Mail = - 1,129.57 + 0.875 # of copies + 76.725 pages per issue

                (-2.83)   (7.35)                (17.2)
```

This relationship is estimated from data on each of 14 issues over the two years and has an adjusted R-square of =0.957. Over this era, JOE averaged 25 pages per issue (ranging from 11 to 51). With seven issues per year, this equation forecasts total printing and mailing costs of $30,019 for 4,000 copies.

JOE became available without charge on a gopher site at Vanderbilt in 1994 and moved to the University of Texas in 1997 (http://www.econ.utexas.edu/joe/) in 1994. The JOE gopher is generating about 25,000 hits per month in 1996 and the subscription list of the printed JOE has dropped to 1,000. The Print & Mail relationship estimated above forecasts a cost of $11,645 for 1,000 copies. The Association will move from a net revenue position of $11,000 ($41,000 - $30,019) in the all print regime to about a zero net ($15,000 - $11,645) with print subscription sales at about a 1,000. Of course, the Association incurs fixed costs in producing JOE that may be similar under both regimes.

[11] The headquarters also publishes a Directory of membership biennially. The Directory became available on-line at the University of Texas in 1995 and is getting about 4,600 hits per month.

Because the Directory comes with membership, we have no measure of the rate of decline in the demand for the print version.

[12] At some point in the future, membership ballots might be solicited and received by the Internet.

[13] The AER's reviewing process is double-blind, with author's names withheld from reviewers and reviewer's names kept from authors. When nearly all working papers are posted on the World Wide Web, the refereeing may become single-blind de facto. Anyone who wants might search the title listing in the working paper file and so identify the author. When working papers are generally accessible on the Net, they would seem to be usable in the editorial process with some saving in cost but with some loss in anonymity.

[14] The fixed costs of a print run (but not typography) would be eliminated entirely if print were abandoned completely. The fixed costs of electronic distribution would replace them in part. Presumably, the more sophisticated the electronic files submitted by authors, the lower the fixed cost of production at the publisher.

[15] Since 1995, the Association has made the JEL available in CD-ROM format instead of print for the same price. The CD-ROM costs about the same to produce on the margin per subscriber as a printed issue of a large journal. The CD-ROM contains the page images of the published journal and is distributed by mail. Its advantage is not reduced cost, but increased subscriber benefit: It adds the power of electronic searching. Therefore, this version is gaining popularity. More than ten percent of the AEA's members opted for the CD-ROM version of JEL in 1996.

[16] The annual meeting contributed a net of about $125,000 in 1995.

[17] Assume the current library subscription rate of $140 yields 20 percent of the AEA's gross and that membership plus ads yields $70, about 40 percent. Assume the shift to electronic distribution lowers total expenditures by 20 percent, a saving of about $140 per library subscription. The campus intranet license then needs to generate double its current amount, about $280.

[18] The notion of doubling the library subscription rate in setting a rate for the campus intranet license is meant to define the Association's probable revenue goals, but not to define the rate structure. The rate structure will need to be tied to something more substantial like enrollment and total research dollars. Alternatively, the rate could be set on the basis of a forecast of the hit rate. OCLC's electronic journal service sets rates on the basis of the number of simultaneous users. The level of rates would likely be set so as to yield about double the current library print subscriptions unless other revenue is forthcoming as discussed in the following paragraphs.

[19] Here is part of the language the AEA prints on the copyright page. "Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or direct commercial advantage and that copies show this notice on the first page or initial screen of a display along with the full citation, including the name of the author."

[20] Jared Sandberg "Cash Advances Aid Electronic Commerce," *Wall Street Journal*,

September 30, 1996 p. B8, reports an offering from CyberCash, a firm working with Visa and several banks. Cybercash put the cost of a transaction at between eight and 31 cents for purchases between $0.25 and $10.

http://www.research.digital.com:80/SRC/millicent/ describes the protocols and tools developed by Digital Equipment Corporation to facilitate Web transactions in fractions of cents. "The key innovations of Millicent are its use of brokers and of scrip. Brokers take care of account management, billing, connection maintenance, and establishing accounts with vendors. Scrip is microcurrency that is only valid within the Millicent-enabled world."

[21] Draft essay at http://alfred.sims.berkeley.edu/iep.html.

[22] See Malcolm Getz, "Petabytes of Information," in *Advances in Library Administration and Organization*, XII (JAI Press, 1994) pp. 203-37. Here are some features that might be added to the network working paper service. Each Association member might receive a private password and encryption key. When the member submits a paper with the password and key, the service would return a time-stamped digital authentication message. This message and the posting would establish ownership to the working paper at the time of submission. The working paper service might include a more elaborate system of tagging papers, including the author's sense of the target audience, degree of originality, sophistication, empirical content, and revision number. The service might include links to comments.

[23] E. Jacquelin Dietz, "The Future of the Journal of Statistics Education," North Carolina State University, mimeo, 1996.

[24] The issue of optimal pricing for three products that share a fixed cost and where cross elasticities are not zero should be explore formally.

[25] David Carpenter and Malcolm Getz, "Evaluation of Library Resources in the Field of Economics: A Case Study," *Collection Management* 20:1/2, 1995, pp. 49-89.

[26] See Malcolm Getz, "Information Storage," *Encyclopedia of Library and Information Science*, Vol. 52, Supplement 15, 1993, pp. 201-39. High density off-site storage might yield an annual cost of $0.30 per volume and so, about $3.00 of capital cost.

[27] OCLC's Electronic Journals Online (EJO) preceded the web-based program. With EJO, OCLC charged publishers for mounting their journals, much as a printers charge for printing. This approach did not attract many publishers. The OCLC website (www.OCLC.org) lists several titles. Here is a sample of subscription rates.

The Online Journal of Current Clinic Trials from Chapman & Hall, distributed by OCLC:

Institutional: $220.00, Individual: $120.00, Student (with ID): $ 49.00, Network (unlimited access): $3,000.00.

Online Journal of Knowledge Synthesis for Nursing from Sigma Theta Tau International, distributed by OCLC: Individuals, $ 60.00; Institutions, $250.00.

[28] OCLC, "Bringing Your Publications Online With OCLC," (Dublin, Ohio, c. 1996) and

OCLC, "A Complete Electronic Journals Solution for Your Library," (Dublin, Ohio, c. 1996).

[29] Malcolm Getz, John J. Siegfried and Kathryn H. Anderson "Adoption of Innovations in Higher Education," *The Quarterly Review of Economics and Finance*, forthcoming.

[30] http://mitpress.mit.edu/jrnls-catalog/snde.html. SNDE is one of six electronic journals offered by the MIT Press in 1996. The library rate includes a license to store the journal on a campus facility and make it available in library reserve services.

[31] http://mitpress.mit.edu/jrnls-catalog/chicago.html puts the subscription rate at $30 for individuals, $125 for libraries, with a $12 fee for downloading an individual article.

••

For additional information about the conference, or The Andrew W. Mellon Foundation's scholarly communication initiatives, please contact Richard Ekman. For additional information about ARL or this web site contact Patricia Brennan, ARL Program Officer at (202) 296-2296.

Return to Office of Scholarly Communication Home Page

ARL Home ARL Scholarly Communication and Technology Home Page

© Association of Research Libraries, Washington, DC
Web Design by Angelo F. Cruz
Maintained by ARL Web Administrator
Last Modified: August 29, 1997

Scholarly Communication and Technology

Conference Organized by The Andrew W. Mellon Foundation

at Emory University
April 24-25, 1997

Copyright © of the papers on this site are held by the individual authors or The Andrew W. Mellon Foundation. Permission is granted to reproduce and distribute copies of these works for nonprofit educational or library purposes, provided that the author, source, and copyright notice are included on each copy. For commercial use, please contact **Richard Ekman** at the The Andrew W. Mellon Foundation.

Session #1 Economics of Electronic Publishing: Cost Issues

Epic: Electronic Publishing is Cheaper

Willis G. Regier, Director
The Johns Hopkins University Press

. . . .

Some time ago the phrase "Electronic publishing is cheaper than print" was promoted from a hope into a credo, taken seriously from Disneyland to Atlanta. "Electronic publishing is cheaper than print" has recently been repeated as often and as faithfully as a mantra but it now looks more and more like a conditional conclusion. It is possible that electronic publishing will eventually be cheaper than print, and it already certain that some types of electronic publishing are cheaper than other types.

In its quest for quality, scholarship likes things that are cheap, loves things that are fast and easy, and worships things that last. The questions nagging us now are how big can media packages get, how long can they last, what impact will they have, who decides, and at what cost. Advocates of inexpensive electronic publishing confront a widespread complaint that there is already an overproduction of scholarship that electronic publishing will make worse. The costs of electronic publishing correlate to a clutch of choices: speeds of access, breadth and depth of content, visibility, flexibility, durability, dependability, definition of community, differentiation, and ease of use. In such a field of choices, there is not a basic cost, or an optimum one, or an upper limit. If cost were no object, the Aeneid would be carved on mountainsides.

Comparative costs guide many crucial decisions in the queasy shift from paper to ether, not only the reproduction costs of print and digitalization, but also the costs of fulfillment, revision, and protection.[1] For the time being, most mainstream digital publications remain dependent on print, either as a publication of record, as with most scholarly journals, or as a center around which electronic sites orbit, as with the Web sites for WIRED, numerous newspapers and magazines, publishers of all stripes, book clubs, and book sellers. In this parallel-publishing environment, print costs remain in place, while the costs of mounting and maintaining a digital presence are added on.

For strategic reasons some publishers have established Web sites with little expectation of recovering those added costs. Washington Times, for example, felt it was necessary to mount a Web site in order to protect its identity as the primary guide to entertainment and restaurants in the nation's capital. The Times cannot risk a rival on the Web that might get a toehold in the market. Similarly, publishers large and small set up Web sites defensively, to maintain an up-to-date profile, to market directly to customers, and to be sure that when and if the Web market matures, they will be ready to compete for it. In the meantime, electronic publishing offers no savings; to the contrary, it requires extra costs that must be recovered somehow. Because these costs are considerable, so is the extra burden on recovery.

The Net is with us and few are so myopic as to think it will go away. But lurking in every discussion of electronic costs is the prospect that print will go away, at least for some forms of publishing. Because on-line publishing vaunts its capability for transferring information speedily, on-line publishers emphasize publishing as nothing else than information transfer. But publishing is not merely the transfer of information: it confers prestige, it competes for attention, it defines a group to itself, sometimes with explicit membership. John Seeley Brown, Director of Research for Xerox, has stressed this community-building function as essential to our comprehension of publishing value.[2] Anyone on a listserve knows that electronic publishing is prone to invasion. The Web exposes a value in print publication that was previously taken for granted: peace.

For a publisher, the costs of electronic publishing can be best understood in the standard publishing sequence: the costs of acquisition, the costs of editing, the costs of the preparation of the first copy, the costs of mass reproduction, the costs of distribution, and the costs of administration. The manufacturing cost of a typical print journal in the humanities consumes about 50% of the journal's operating budget, and shipping and warehousing can eat up another 10%; to a cybershark these percentages look like chunks of fat removable in big bites. Those who confidently declare that electronic publishing is cheaper than print focus chiefly on perceived savings in reproduction and distribution, on the premise that once the first copy is prepared its reproduction and transmission circumvent the costs of printing, paper, ink, packaging, shipping, spoilage, and inventory.

This premise is the pot of gold we all pursue, but experience has shown that there are at least three holes in its rainbow. First, electronic publishing adds numerous new costs to preparation of the first copy. Second, the savings enjoyed by the publisher are made possible only if the end user, whether a library or an individual, has also invested a hefty sum in making it possible to receive the publication. And third, both the scholarly publisher and the end-user alike are dependent upon even greater costs being born by universities and their libraries.

As costs became gradually more predictable for Project Muse, Marie Hansen calculated that the additional costs for preparing parallel print and electronic journals is about 130% of the cost of print only. Even if print versions were dropped, the costs to produce the first copy ready for mounting on a server would be as high as 90% of the cost of a paper journal.[3] The cost savings for printing, storage, shipping, and spoilage are substantial, but in the digital realm they are replaced by new costs: system administration, content cataloging, tagging, translating codes, checking them, inserting links, checking them, network charges, computer and peripherals charges, and additional customer service.

In the near term there are also high costs for acquisitions. It has taken longer than expected to negotiate contracts with journal sponsors, to obtain permissions, and to acclimate journal editors to the steps required for realizing the efficiencies of the digital environment. Electronic editors play fast and loose with copyright, always waving the banner of "fair use" while blithely removing copyright notices from texts and images. Piracy is not only a foreign problem: it occurs everyday in name of freedom. Explaining to electronic editors why copyright is in their best interests, and thus worthy of observance, has been just one time-consuming task. As Project Muse enters its third year, we see more clearly the costs of rearing an infant.

The Supra of the Infra

The fundamental costs of a university infrastructure are enormous. The Homewood campus at Hopkins is home to 5200 students, faculty, and staff who want connections to the Net. The start-up costs for rewiring the campus to UTP--at a rate of about $150 per connection--would have been impossibly high for the University if not for $1 million in help from the Pew Trust. According to Bill Winn, the Associate Director for Academic Computing at the Homewood campus of Hopkins, it costs $20 per month for each person to connect to the campus network. The network itself costs $1 million per year to maintain, and an additional $200,000 to support PPP (point-to-point protocol) connections. The annual bill to provide Net access to the 900 students who live off-campus is $200,000. The fee to the campus's Internet Service Provider for a 4-megabit-per-second Net-link, plus maintenance and management, costs the University about $50,000 per year.

Students, he says, require high maintenance: if their connections are insecure, it's often because they've been ripped from the wall. Last year students in engineering attempted to install a software upgrade for a switch that exceeded their wildest dreams: it shut down the University's system for more than a week. If you're counting, that adds up to about $20,000 of lost Net access, not to mention the costs of repair.

In 1996, Johns Hopkins University budgeted $70,000 for hardware maintenance and $175,000 for hardware upgrades, chiefly to handle rapidly increasing traffic. The million-dollar budget supports a staff of three technicians, an engineer, a software analyst, and a director for networking who are so busy handling day-to-day problems and requests that it is clear to most people on campus that additional staff is needed. Their overhead is kept to a minimum, since some key people are based in a trailer parked below the Milton Eisenhower Library.

Bill Winn believes that the $20-per-month access charge is comparable to other campuses elsewhere in the United States, a useful point of departure for all other cost estimates. When it costs $240 a year per person to link a computer to the Net, the University's administration confronts a cost chasm. This is $240 that cannot be spent on something else. And this is only a chip of the iceberg: each department bears most of the costs for its own infrastructure, at costs unexpected in the days of typewriters and paper memos.

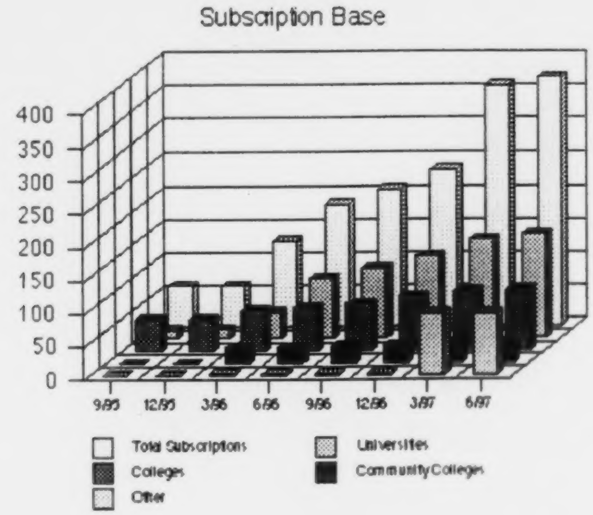
Further, in order to make this initial investment worthwhile, more expensive investments must continue to be made: upgrades, peripherals, database access fees, consultants, and specialized software. It is no wonder that the virtual bloom has fled the virtual rose for many colleges, who have second thoughts about their level of commitment to Net access. To some extent, electronic publishers are still stymied by the lag between the Net's ability to produce and its readers' ability to receive. That lag bears a price tag, and some institutions cannot or will not pay, most state governments cannot pick up the bill, and the federal government is increasingly reluctant to reserve space or investment for scholarly networking.

Optimum Optimism

Every sidewalk philosopher has speculated whether electronic publishing will exacerbate monopolies and class divisions, or whether a slow, steady spread of access will lower costs and lead to greater democratization. The Net is full of threads on the inconsistent costs of Net access from place to place.[4] Depending on my morning caffeine intake, I am more or less optimistic about the liberating prospects offered by the Net for our era. It may be that computers will be as ubiquitous as TVs and a Net connection as cheap as a telephone.[5] But for now, when I focus on the role of the Net in higher education, I usually see higher costs, and foresee only more and more differentiation based upon costs and the ability to recover them. What must not be lost in these sober comparisons is that the conversion from print to pixels is not merely a change of clothes: it is an enormous expansion of capability.

Added costs purchase substantial added value. Under the domain plan that Muse, Jstor, Artfl, and other experiments are refining, we have already accomplished no less than seven olympic leaps in scholarly transmission. Here is the hallowed litany: (1) instead of a library maintaining one copy of a work that can be read by one person at one time, the work can now be read by an entire campus simultaneously; (2) instead of having to search for a location and hope that a work is not checked out or misshelved, a user can find the full text at the instant it is identified; (3) the work can be read in the context of a large and extensible congregation of journals, including back issues, each as easily accessible as the first; (4) the work is capable of being transformed without disturbing an original copy; pages can be copied without being ripped out; students can make copies without complaining that the photocopier is jammed or out of toner; (5) the work can be electronically searched; (6) there is no worry about misplacing the work or returning it by a due date; and (7) the increase in costs, if honestly reflected by a corresponding increase in price, permit libraries to spend a little more to be able to offer a lot more content, expanding their holdings geometrically while increasing their costs arithmetically. This is not just pie in the sky: our readers are reaping real fruit. Project Muse has already attracted 100 library subscribers who previously subscribed to none of our print journals, including libraries in museums and community colleges. (See Graph 1.)

Graph 1



Even if some claims for the digital revolution are ridiculously inflated, its agents can confidently claim that the revolution has occurred with unprecedented self-consciousness and organizational care. That care deserves a few choruses of praise. I am thankful for the assistance of commercial presses for their support for standardization, their defense of copyright, their vigilance against piracy, and their scrutiny of current and pending legislation. I am thankful for the frank and frequent discussions between publishers and librarians. Conversations with Jim Neal often remind me of a home truth: libraries are the original multimedium. For multiple reasons librarians' reactions to the systemic costs of digitalization are immediately relevant to publishing decisions. I am thankful for the support of the Mellon Foundation. Its key role in the development of digital scholarly communication has not only saved universities delay, risk, and anxiety, but has put the universities where they can do the most good: out in front, experimenting, thinking things through. If not for the Mellon Foundation and its projects the growth of the Net would shuttle between large corporations and isolated individuals, with maddening secrecy and without much interest in the special needs of scholarship and the special costs it encumbers. Efforts to create a cheaper and more attractive home for STM studies would have stuck to the starting blocks. Libraries would be asked to acquire extraordinarily expensive databases without a clue about the relationship between price and actual costs. If the digital revolution is a revolution rather than a colossal marketing scheme, it is because so many people and institutions are involved and invested.

For Muse the greatest cost is for personnel. For decades, it has been possible to maintain a journals program staffed by literate and dedicated people; Muse employees also have to well beyond computer literacy and masters of complex skills. To raise Muse from infancy, they must also be virtually parental -- creative, patient, resourceful, and endowed with heroic stamina. Because their jobs require higher and higher levels of education and technical skill, starting positions are more expensive. Disregarding administrative costs, the staff of Muse cost about 20% more per capita per month than the staff of print journals.

We are just beginning to understand the costs of hiring, training, and retaining qualified staff. Because the skills of the Project Muse team are pioneering, those who succeed are much in demand, and are subject to recruitment raiding for still higher salaries. Due to the inordinate pressures put upon them--not only the stress of the grant schedule, the frustrations of downtime, and the frictions of incompatible programming--but also their anxiety about their futures and the very real resentment projected by the staff in print journals, these young people may grow old at a rate faster than Bill Gates can update software. The next time a rosy-cheeked cherub cheerfully announces the death of print, let him look into the bloodshot eyes of the Muse staff. What seemed to be freshness and precocity a couple of years ago now shows signs of premature burn out.

Excluding independent contractor costs, personnel costs account for 46% of the start-up and maintenance costs for Project Muse. Including independent contractor costs, which are themselves chiefly a matter of personnel, that percentage rises to 59%.

Second only to personnel, the largest expense has been hardware, accounting for 12% of total costs.[6] Third is rent, at 3.3%. Fourth, surprisingly, has been travel, requiring 2.9% of investment. The travel budget is a direct reflection of the extensive need to negotiate on every frontier: with the learned societies and editorial boards that run the journals, with the librarians who buy them, and with editors who want to move their journals to Muse. In the first two years of Muse's development, our efforts to build Muse were distracted by the novelties of the Net--training staff, dealing with journal sponsors, conversing with libraries--each a task as vital as the selection of software or the conversion of codes. Marketing was kept to a minimum until we had a complete package to deliver. With the completion of the forty-journal base last December, we are now in high gear marketing Muse, so marketing expenses will begin to affect all percentages. Travel and exhibits will have still higher costs as we strive to attract a subscription base strong enough to make Muse self-supporting.

The Electronic Market

Marketing on the Web is a different creature than marketing via print or radio, because it must contend both with misinformation and with the difficulty of finding an audience. Misinformation about an electronic site shows up in the same search that finds the site itself and may require quick response. Muse responds readily enough to the Net's search engines, but only if the person is searching. Even then, the searcher can only read text if the searcher's library has already subscribed. At the December 1996 Modern Language Association exhibit, about half of the persons who expressed their wish that they could subscribe to Muse belonged to universities that already did, but the scholars didn't know it. With usage data looming as a subscription criterion, we will cannot rest after a subscription is sold; we still have to reach the end user.

The marketplace itself is changing. Most conspicuously, the unexpected formation of library consortia has reshaped many a business plan. Expectations of library sales have often hung fire while libraries consorted, but in the long run it is likely that by stimulating these consortia, electronic publishing will have served an important catalytic function for discovering and implementing many kinds of efficiencies.

The Net Market is enormous and enormously fragmented.[7] In the next year there will be numerous marketing experiments on the Web. New and improved tools emerge every month that will help us reply to scholars with specific requests, complaints, and inquiries. Publishers are cautiously optimistic that electronic marketing will prove more advantageous than bulk mail, and it will certainly be cheaper. Already most university presses have their catalogs on line and many are establishing on-line ordering services.

Customer service is another high cost--at present, much higher than for print journals. Today it takes one customer service agent to attend to 400 Project Muse subscriptions, while a customer service agent for print journals manages about 10,000 subscriptions. But the future offers bright hope. In February, our customer service agent for Project Muse sent an e-mail message to 39 past-due subscribers to Muse who were not with a consortium. Within 24 hours of sending this letter, she received 29 responses to it, and four more arrived the next day. Each thanked her for sending the letter, and all 33 renewed for 1997. Here the advantages of on-line communication are obvious and immediate.

There are also costs that are difficult or impossible to track or quantify, like intellectual costs. It is these costs that have emerged as the next vexed problem in the development of electronic scholarly resources. The problem has three prongs.

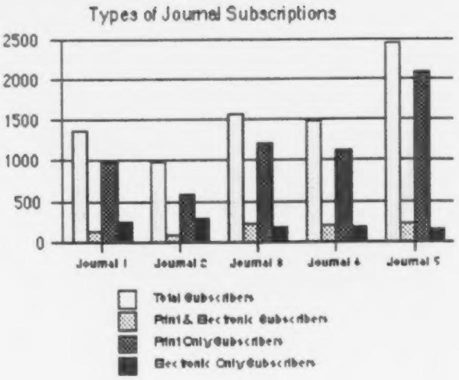
One is scholarly skepticism about the value of electronic publishing for tenure and promotion.

Another is the fluidity of the Web, which for all its nautical metaphors often seems a murky flood. Journal editors are anxious about the futures of their journals and hesitant about entrusting them to a medium as fleeting as electricity. Well aware of past losses, scholarship generally prefers the medium most likely to last. This preference is firmly based: some ideas take time to hatch, some messages take years to sprout, and the gush and backwash of the Web seem unstable or engulfing. Scholars care that their work endures; that it is a heritage; that if they care for it well it will live longer than they do. Scholars who know and use the Net often encounter defunct URLs, obsolete references, wretched writing, Web sites that bloomed like gardenias and softened into mulch, and mistakes of every kind. Ephemerae appear more ephemeral on screen. Chief among the concerns expressed by librarians interested in purchasing electronic publications is whether the publication is likely to be around next year and the year after.

The third prong is the sharpest: will electronic publishing be able to recover the operating costs of scholarship, the costs of editing, of maintaining a membership, and of defending a niche in the pantheon? If journals are to migrate to electronic formats, they will have to be able to survive there, and survive the transition, too: the current competition is part endurance, part sprint. Since parallel publishing in print and on line costs more, library budgets will either have to pay more to sustain dual-format journals, or cut them, or cut other journals to support them.

In the short term, at least, there is reassurance in numbers. Rather than erode reader and subscription base, electronic versions of journals actually increase them. (See graph 2). Even if paper subscriptions dwindle, it appears that the increase in subscriptions and readership will last. Of course, means for cost recovery for each journal must also last, which is why different publishers are trying different pricing strategies.

Graph 2



Competition in the electronic environment is expensive and aggressive (a favorite book for Netizens is Sun Tzu's Art of War).[8] Foundation assistance can enable university presses and libraries to enter the competition, but it is uncertain whether their efforts can compete for very long when foundation support ends. Scholarship has deep reservoirs of learning and good will, but next to no savings; one bad year could wipe out a hundred-year-old journal. Unless journal publishers and editors can migrate quickly and establish a system to recover costs successfully, the razzle-dazzle of paper-thin monitors will cover a casualty list as thick as a tomb.

This risk has shifted attention from the costs of production and distribution to the costs of acquisition. Publishers and their partners are trying to determine what costs must be paid to attract scholars to contribute to their sites. It is obvious that a moment after a scholar has completed a work, a few more keystrokes can put the work on the Web without bothering a publisher, librarian, faculty committee, or foundation officer. Indeed, electronic publishing is cheaper than print, if you rule out development, refereeing, editing, design, coding, updating, marketing, accounting, and interlinking. Further, there are numerous scholars who believe they should be well paid for their scholarship or their editing. Stipends paid by commercial publishers have raised their editors' financial expectations, which in turn exacerbated the current crisis in STM journals. Retention of such stipends will devour savings otherwise achieved by digitalization.

What is now at issue is what each added value is worth. Competitive programs are now testing the academic market to see how much it wants and how much it will pay, whether page images are preferable to HTML, whether pricing should sequester electronic versions or bundle them into to an omnibus price, what degree of cataloging and linking and tagging are desired, what screen features make sense, and a realm of other differentia, not least of which is the filtering of the true from the spew. We expect to see significant differences between the costs and prices of scientific and humanities journals, and with our library partners scrutinizing real usage and comparative costs, we expect these differences will be less and less defensible. And we expect to see gradual but salutary changes in scholarship itself as different disciplines come to terms with the high visibility of electronic media. Ballooning literature surveys, for instance, are prime candidates for reform. We expect to see a clearer separation of reputation, with all that a reputation is worth, as professionally managed electronic media distance their offerings from the Web sites of hobbyists, amateurs, and cranks. Finally, we expect to see shifts in academic collaboration and shifts within disciplines. As electronic publishing increases its pressure on hiring, evaluation, tenure, and promotion, the certification and prestige functions of publishers will increasingly depend on their attention to the emerging criteria of e-publishing, in which costs are measured against benefits that print could never offer.

NOTES:

[1] Piracy is a real threat. According to the Software Publishers Association, about $13 billion in sales was lost due to piracy in 1996. See the SPA homepage against piracy: http://www.spa.org/piracy/homepage.htm.

[2] See John Seely Brown and Paul Duguid, "The Social Life of Documents," Release 1.0, October 1995. See http://www.edventure.com/release1/abstracts/9510.html.

[3] Marie Hansen, "Pricing Issues for Electronic Journals," unpublished.

[4] For example, http://www2.iphil.net/ph-isp/1995- Dec/0393.html.

[5] An article in Upside forecast that Internet customer services could save businesses 25% to 50% of the cost of traditional telephonic customer support. David Kline, "Reshaping the way America Does Business," Upside Online, August 5,1996.

[6] For hardware specifications for each member of the Project Muse staff, see: http://calliope.jhu.edu/poj-descrip/tech_specs.html.

[7] There is also enormous disagreement about how enormous it is. Recent estimates vary between 5.8 million and 35 million users. See http://www.cyberatlas.com/market.html.

[8] See, for instance: http://www.pic.net/%7Ewlevinso/sun_tzu.html; http://www.geocities.com/Athens/4884/; and http://www.kimsoft.com/polwar.htm.

……

For additional information about the conference, or The Andrew W. Mellon Foundation's scholarly communication initiatives, please contact Richard Ekman. For additional information about ARL or this web site contact Patricia Brennan, ARL Program Officer at (202) 296-2296.

Return to Office of Scholarly Communication Home Page

ARL Home ARL Scholarly Communication and Technology Home Page

© Association of Research Libraries, Washington, DC

Web Design by Angelo F. Cruz
Maintained by ARL Web Administrator
Last Modified: August 29, 1997

Scholarly Communication and Technology

Conference Organized by The Andrew W. Mellon Foundation

at Emory University
April 24-25, 1997

Copyright © of the papers on this site are held by the individual authors or The Andrew W. Mellon Foundation. Permission is granted to reproduce and distribute copies of these works for nonprofit educational or library purposes, provided that the author, source, and copyright notice are included on each copy. For commercial use, please contact Richard Ekman at the The Andrew W. Mellon Foundation.

Session #1 Economics of Electronic Publishing: Cost Issues

The Use of Electronic Scholarly Journals

Models of Analysis and Data Drawn from the Project Muse Experience at Johns Hopkins University

James G. Neal
Sheridan Director
Milton S. Eisenhower Library
The Johns Hopkins University
Baltimore, Maryland 21218
jneal@jhu.edu

• • • • • • • • • • • • • • • • • • • •

This paper outlines a series of quantitative and qualitative models for understanding and evaluating the use of electronic scholarly journals, and summarizes data based on the experience of Project Muse at Johns Hopkins University and early feedback received from subscribing libraries.

Project Muse is a collaborative initiative between the Press and the libraries at Johns Hopkins University to provide network-based access to scholarly journals including titles in the humanities, social sciences and mathematics. Launched with electronic versions of forty titles still published in print, Project Muse coverage has now been expanded to include electronic-only publications. Funded initially by grants from the Mellon Foundation and the National Endowment for the Humanities, Project Muse seeks to create a successful model for electronic scholarly publishing characterized by affordability and wide availability. It has been designed to take advantage of new technical capabilities in the creation and storage of electronic documents. It has been developed to provide a range of subscription options for individual libraries and consortiums. It is based on a very liberal use and re-use approach that encourages any non-commercial activity within the bounds of the subscribing organization.

Project Muse has been produced from the outset for usability, with a focus on user-centered features. This has evolved as a participative and interactive process, soliciting input and feedback from users, and integrating user guidance components into the system. An online survey is available to all users and libraries are providing information about the local implementation and the results of campus and community focus group discussions on Project Muse. As the number of subscribing libraries expands and the activity grows, a valuable database of user experiences, attitudes and behaviors will accumulate. A new feature will be the ability to track and analyze individual search sessions and to observe closely user activities. This will enable monitoring the impact of new capabilities and the efficiency of searching practices.

Six models of use analysis are discussed in this paper which cover both the macro or library-level and the micro or individual user-level activity:

1. subscribing organizations - which libraries are subscribing to Project Muse and how do they compare with the base of print journal customers

2. subscriber behaviors - how do libraries respond as access to electronic journals is introduced and expanded, and in particular, how are acquisitions like Project Muse accommodated in service and collection development programs and budgets

3. user demography - what are the characteristics of the individual user population, in such areas as status, background/experience, motivation, attitudes and expectations

4. user behaviors - how do individuals respond to the availability of scholarly materials in electronic format as they explore the capabilities of the system and execute requests for information

5. user satisfaction - what objectives do users bring to network-based access to scholarly information, and how do users evaluate system design and performance and the quality of search results

6. user impact - how are user research and information-seeking activities being shaped by access to full-text journal databases like Project Muse

One of the objectives of Project Muse is to achieve full cost recovery status by the completion of the grant funding period in 1998. Therefore, it is important to monitor the growth in the base of subscribing libraries and to evaluate the impact on the print journal business of the Press. An analysis of those libraries subscribing to the full Project Muse database as of June 1997 (approximately 400 libraries) demonstrates a very significant expansion in the college, community college and now public library settings with very low or no history of subscriptions to the print journals. The result is a noteworthy expansion in access to Hopkins Press titles with 70 percent of the subscribing libraries currently purchasing less than 50 percent of the titles in print, and over one-fourth acquiring no print journals from the Hopkins Press.

PROJECT MUSE
SUBSCRIBING LIBRARIES

ARL Universities	-	65 libraries
Other Universities	-	128 libraries
Liberal Arts Colleges	-	101 libraries
Community Colleges	-	53 libraries
Public Libraries	-	3 library systems

PROJECT MUSE
CUSTOMER PRINT SUBSCRIPTIONS

No. of Print Subscriptions	Percentage of Libraries
0	27.8
1-4	7.6
5-9	6.8
10-14	15.8
15-19	13.0
20-24	11.8
25-29	9.0
30-34	6.2
35-40	2.0

One of the explanations for these patterns of subscription activity is the purchase arrangement for Project Muse. Over 90 percent of the libraries are subscribing to the full Project Muse database of 42 titles. And due to very favorable group purchase rates, nearly 80 percent of Project Muse subscribers are part of consortial contracts. The cooperative approach to providing access to electronic databases by libraries in a state or region is widely documented, and the Project Muse experience further evidences this phenomenon.

Another objective of Project Muse is to enable libraries to understand the use of collections and thus to make informed acquisitions and retention decisions. The impact on collection development behaviors will be critical, as libraries do indicate intentions to cancel print duplicates of Muse titles and to monitor carefully the information provided on individual electronic title and article activity. Use information is beginning to flow to subscribing libraries, but there is no evidence yet of journal cancellations for Hopkins Press titles.

An important area of analysis is user demography, that is the characteristics of the individuals searching the Project Muse database. An online user survey and focus group discussions are beginning to provide some insights:

> -- The status of the user, that is undergraduate student, graduate student, faculty, staff, community member, or library employee. As Project Muse is introduced, library staff are typically the heaviest users, followed by a growth in student use as campus awareness and understanding expands.

-- Type of institution, that is research university, comprehensive university, liberal arts college, community college, or public library setting. As Project Muse subscriptions have increased and access has extended into new campus settings, there has initially been heavier use in the research universities and liberal arts colleges where there is either traditional awareness of Project Muse titles or organized and successful programs to promote availability.

-- The computer experience of users, that is familiarity with searching full-text electronic databases through a Web interface. Project Muse users tend to be knowledgeable Internet searchers who have significant comfort with Web browsers, graphical presentations of information, and constructing searches in textual files.

-- The location of use, that is in-library, on-campus in faculty office and student residence hall, or off-campus. Preliminary data indicates that the searching of Project Muse is taking place predominantly on library-based equipment. This can be explained by the inadequate network infrastructure that persists at many campuses or the general lack of awareness of Project Muse until a user is informed by library staff about its availability during a reference exchange.

-- The browsers used to search the Project Muse database. An analysis of searches over an 18-month period confirms that Netscape browsers are used now in over 98 percent of the database activity, with a declining percentage of Lynx and other non-graphical options.

Project Muse enables searching by author, title, or keyword, in the table of contents or full-text of the journals, and across all the journals or just selected titles. All articles are indexed with Library of Congress subject headings. Hypertext links in table of contents, articles, citations, endnotes, author bibliographies, and illustrations allow efficient navigation of the database. User searching behavior is an important area for investigation, and some preliminary trends can be identified:

-- The predominate search strategy is keyword, with author and title inquiries occurring much less frequently. This can be partially explained by the heavy undergraduate student use of the database and the rich results enabled by keyword strategies.

-- Use of the database is equally distributed across the primary content elements: table of contents, article abstracts, images linked to text, and the articles. An issue for future analysis is the movement of users among these files.

-- Given the substantial investment in the creation of LC subject headings and the maintenance of a structured thesaurus to enhance access to articles, their value to search results and user success is being monitored carefully.

-- With the expansion of both internal and external hypertext links, the power of the Web searching environment is being observed, the user productivity gains are being monitored, and the willingness to navigate in an electronic journal database is being tested.

-- Users are directed to the Project Muse database through several channels. Libraries are providing links from the bibliographic record for titles in the online catalog. Library Web sites highlight Project Muse or collections of electronic journals. Subject pages list the

Project Muse titles that cluster in a particular discipline.

-- Users are made aware of Project Muse through a variety of promotional and educational strategies. Brochures and point-of-use information are being prepared. In some cases, campus media have included descriptive articles. Library instructional efforts have focused on Project Muse and its structure and searching capabilities.

-- Printing and downloading to disk are important services linked to the effective use of Project Muse, given the general unwillingness of users to read articles online. Libraries have an interest in maximizing turnover on limited computer equipment, and are focused on implementing cost-recovery printing programs.

-- Project Muse is increasingly enabling users to communicate with publishers, journal editors, and the authors of articles through e-mail links embedded in the database. Correspondence has been at a very low level, but is projected to expand as graduate student and faculty use increases and familiarity and comfort with this feature expands.

With other 400 subscribing libraries and over three million potential users of Project Muse in the communities served, it is possible to document global use trends and the changing intensity of searching activity:

```
                     PROJECT MUSE
                  GLOBAL USE TRENDS

       4th Quarter 1996               1st Quarter 1997

   1,833,692  Requests               2,618,069  (+42.8%)
      19,922  Per Day                   29,090  (+46.0%)
         199  Subscribers                  322  (+61.8%)
       9,214  Per Subscriber             8,112  (-12.0%)
```

The progression of use over time as a library introduces access to Project Muse is being monitored. Early analysis suggests that the first two quarters of availability produce low levels of use, while third quarter use expands significantly.

User satisfaction with the quality and effectiveness of Project Muse will be the central factor in its long-term success. Interactions with users seek to understand expectations, response to system design and performance, and satisfaction with results. The degree to which individuals and libraries are taking advantage of expansive fair use capabilities should also be gauged.

Project Muse has focused on various technical considerations to maximize the dependability and efficiency of user searching. Detailed information on platforms and browsers is collected, for example, and access denials and other server responses which might indicate errors are automatically logged and routed for staff investigation.

Expectations for technology are generally consistent: more content, expanded access, greater convenience, new capabilities, cost reduction, and enhanced productivity. It will be important to monitor the impact of Project Muse in the subscribing communities and to assess whether it is delivering a positive and effective experience for users.

Over the now thirty-plus years of library automation activities, we have learned about those conditions which improve the positive impact of technology, and Project Muse and its implementation must respond to these needs:

-- if the computer package is more decentralized so that is in the hands of users (control);

-- if more developed computing capacity is available (power);

-- if users have greater competency and experience with computing (training);

-- if the developers are more responsive to the expressed needs of the user regarding design and operation of systems (support); and

-- if users routinely rather than selectively use computers and networked-based information systems (opportunity).

A recent study carried out jointly by the Association of Research Libraries and the Association of American Universities on electronic scholarly publishing identified a series of critical performance attributes for technology: ease of use, timeliness, responsiveness, accuracy, authenticity, predictability, adaptability, relevance, eligibility, cost, recovery, innovation and extensibility. These qualitative characteristics are essential benchmarks for evaluating Project Muse.

Charles Hildreth, an early investigator of online library catalog systems, established a series of analyses which serve as well our review of full-text databases. Hildreth cited five components for understanding the user interface:

-- physical, including the input-output equipment, and the structure and location of the workstation;

-- organizational, including the institutional setting, the availability of staff assistance, and the provision of user aids;

-- personal, including the abilities, experience, objectives and needs of the user;

-- communications, including the language and techniques of interaction; and

-- functional, including control of operations, search formulation and output.

Hildreth views user support in terms of five system qualities: easy to use, friendly and cordial, protective and forgiving, reliable and responsive, and adaptive and flexible. These various elements can be summarized in terms of three general characteristics: audience suitability or the degree to which effective use of the system is self-explanatory; metaphorical consistency or the extent to which a logical framework is provided; and display legibility. Systems which strive to support the user reflect a concern for these elements and emphasize both simplicity of design and searching power.

It is also important to maximize the core advantages of using information in digital formats:

-- accessibility, that is delivery to locations wherever users can obtain network connections

-- searchibility, that is the range of strategies that can be used to draw relevant information out of the database

-- currency, that is the ability to make publications available much earlier than is possible for print versions

-- researchability, that is the posing of questions in the digital environment that could not even be conceived with print materials

-- interdisciplinarity, that is the ability to conduct inquiries across publications in a range of diverse disciplines and discover new but related information

-- multimedia, that is access to text, sound, images, video in an integrated presentation

-- linkability, that is the hypertext connections that can be established among diverse and remote information sources

-- interactive, that is the enhancement of user control and influence over the flow of information and the communication that can be integrated into the searching activity

Project Muse will be evaluated against these quantitative and qualitative models. It's success will ultimately be determined by its support for the electronic scholarly publishing objectives outlined in the ARL/AAU work:

-- foster a competitive market for scholarly publishing by providing realistic alternatives to prevailing commercial publishing options

-- develop policies for intellectual property management emphasizing broad and easy distribution and reuse of material

-- encourage innovative applications of information technology to enrich and expand the means for distributing research and scholarship

-- assure that new channels of scholarly communication sustain quality requirements and contribute to promotion and tenure processes

-- enable the permanent archiving of research publications and scholarly communication in digital formats

....

For additional information about the conference, or The Andrew W. Mellon Foundation's scholarly communication initiatives, please contact Richard Ekman. For additional information about ARL or this web site contact Patricia Brennan, ARL Program Officer at (202) 296-2296.

Return to Office of Scholarly Communication Home Page

ARL Home ARL Scholarly Communication and Technology Home Page

© Association of Research Libraries, Washington, DC
Web Design by *Angelo F. Cruz*
Maintained by *ARL Web Administrator*
Last Modified: August 29, 1997

Scholarly Communication and Technology

Conference Organized by The Andrew W. Mellon Foundation

at Emory University
April 24-25, 1997

Copyright © of the papers on this site are held by the individual authors or The Andrew W. Mellon Foundation. Permission is granted to reproduce and distribute copies of these works for nonprofit educational or library purposes, provided that the author, source, and copyright notice are included on each copy. For commercial use, please contact Richard Ekman at the The Andrew W. Mellon Foundation.

Session #1 Economics of Electronic Publishing: Cost Issues

The Library and the University Press: Two Views of the Costs and Problems of the Current System of Scholarly Publishing

Sandra Whisler
Assistant Director, Electronic Publishing
University of California Press

Susan F. Rosenblatt
Deputy University Librarian
University of California at Berkeley

• •

ABSTRACT:

The costs of scholarly publishing have become unsustainable for both research libraries and university presses. This paper discusses how the transition to electronic journal publishing changes the ways in which these two participants in the scholarly communication process begin to analyze and attempt to control their cost structures in order to remain economically viable. During the near-term future, pressure to maintain both print and electronic dissemination will be great. Libraries and their users will be reluctant to abandon a known archival format, and capital

investments in the technical infrastructure needed to deliver scholarly information electronically may be made slowly. For publishers, the need to cover first copy costs and to continue serving a market demand for print will create a significant transitional period during which both print and electronic formats must be produced and funded. Moreover, the transition to fully electronic publication, although likely to reduce operational costs for libraries slightly in the short run and significantly in the long run, creates very serious potential revenue interruptions for presses. To ensure fiscal stability during an indeterminate transition phase, many publishers have proposed pricing models for electronic journals that are based on existing print subscription prices and that include multi-year guarantees of price adjustments to cover both inflation and expansion in the content offered. Although the rates of these price adjustments are frequently lower than anticipated for print subscriptions, they are greater than the expected increases to libraries' budgets for collections. Therefore, libraries, whose historical funding models for collections lack adjustments adequate to compensate for actual inflation, are caught in the dilemma posed by many publishers' current pricing structures for electronic journals: the offer of a multi-year reduction in the rate of inflation for high-value commercial journals is attractive when compared to the anticipated inflation in print journals; yet accepting that model would protect a rising share of library collection budgets for high-inflation journals which would then rapidly crowd out other scholarly publications. The short-term measures that the library and press individually might rationally employ to maintain fiscal stability may have far reaching negative implications for the economic viability of the system of scholarly communication as a whole, particularly for university presses.

INTRODUCTION:

The crisis in scholarly communication has been well-known for almost two decades. In a statement that could be written today, Patricia Battin wrote in 1982:

> During the decade of the 1970's, librarians faced declining budgets, increasing volume of publication, relentless inflation, space constraints, soaring labor costs, a horrifying recognition of the enormous preservation problems in our stacks, increasing devastation of our collections by both casual and professional theft, and continuing pressure from scholars for rapid access to a growing body of literature. It is ironic that both librarians and publishers introduced computer applications into libraries and publishing houses to save the book, not to replace it. Both were looking for ways to reduce labor costs rather than for visionary attempts to redefine the process of scholarly communication. . . . The former coalition shattered and publishers, scholars and librarians became adversaries in a new and unprecedented struggle to survive in the new environment, each trying in his or her own way to preserve the past and each seeing the other as adversary to that objective.[1]

LIBRARY COSTS

Library Materials:

Print:

The results of the economic crisis in the system of scholarly publishing were documented statistically for the first time in University Libraries and Scholarly Communication.[2] Some of the principal findings included the facts that although materials and binding expenditures remained a relatively constant percentage of total library expenses, there had been a hidden, but significant, change in the ratio of books and serials expenses; and that although materials expenditures had steadily risen, the average annual numbers of volumes added to library collections continued to decline. Not only were libraries spending more and receiving fewer items in absolute terms, libraries were also collecting an ever smaller percentage of the world's annual output of scholarly publications; from 1974, even increases in university press outputs outstripped library acquisition rate increases.

Moreover, the study documented that certain fields experiencing some of the greatest increases in their share of the total output were precisely those with the highest average per-volume hardcover prices: business, law, medicine, and technology. According to the report, science had the highest average prices and remained at a more or less constant and significant market share of about 9.5 percent; titles in arts and humanities, social sciences, and business experienced price increase rates closer to the GNP deflator (p.xix).

Another finding was that serials prices consistently increased faster than inflation, experiencing an overall annual inflation rate of more than 11 percent from 1986 to 1990. Prices of scientific and technical journals rose at the highest rates (13.5 percent per year, on average from 1970 to 1990), and the most expensive serials experienced the largest relative price increases. In contrast, book prices inflated at 7.2 percent per year, while average general annual inflation was approximately 6.1 percent. The report suggests that in certain institutions, science journals could comprise only 29 percent of the total number of journal subscriptions yet consume as much as 65 percent of the serials budget. According to the report, "three European commercial publishers (Elsevier, Pergamon, and Springer . . .) accounted for 43 percent of the increase in serials expenditures at one university between 1986 and 1987" (p.xxi). The report does not introduce the question of the extent to which these inflation rates in the prices of scientific journals reflect increasing costs of production, expansion in content, the market value of the information itself -- a value that might extend well beyond the university, or price gouging.

In 1996, Brian Hawkins updated the study and found the following:

> In the 15-year period from 1981 to 1995, the library acquisition budgets of 89 of the nation's finest schools nearly tripled, and in real dollars increased by an average of 82 % when corrected for inflation, using the Consumer Price Index (CPI) . . . the average library in this elite group of libraries lost 38% of its buying power during this period . . . In those 15 years, the inflation rate for acquisitions was consistently in the mid teens. Although the costs of . . . monographs did not rise quite as fast, the cost of some serials -- especially those in the sciences -- increased over 20% a year. If these trends continue, by the year 2030 the acquisitions budgets of our finest libraries will have only 20% of the buying power they had just 50 years earlier . . . As dire as these projections may be, it should be recognized that they are based on the precarious assumption that library acquisitions will increase an average of 8% compounded per year as they have for the past 15 years. This amount is nearly three times inflation, and nearly twice the amount of total increases in the cost of higher education.[3]

However, Hawkins notes that the trend line for average increases in Library acquisition budgets is downward. While average acquisition budget increases were 9.67 percent during 1981-85, the increases were only 5.4 percent during 1991-95,. Hawkins extrapolates from these figures to conclude that if inflation in the price of scholarly information were to remain steady and library acquisitions budgets to increase at a rate similar to that of 1991-95, then libraries would have only 20 percent of their 1981 purchasing power by 2007.

Other analyses lead to similar conclusions.

Harrassowitz regularly alerts libraries to subscription pricing information so that its customers can plan in advance to adjust purchasing patterns to stay within budget. In November 1996, Harrassowitz provided firm 1997/98 subscription pricing for six publishers publishing the majority of the STM journals.[4] The announced price increases ranged from 1.2 percent to 22 percent, averaging 11.15 percent. A weighted average based on the numbers of titles published by the publishers yields an average of 11.82 percent. Harrassowitz further provided an analysis of the impact of announced price increases on particular types of libraries. According to Harrassowitz, those libraries categorized as General Academic/including Sci-Tech can expect price increases from the six publishers ranging from 6.6 percent to 22.4 percent, with an average increase of approximately 13.87 percent.

An interesting discussion of the problem from the point of view of one scientific library has been prepared by Peter Brueggeman, Head of the Scripps Institution of Oceanography (SIO) Library at UCSD.[5] At SIO, journal subscription prices inflated 57 percent in the five years from 1992 to 1996; the average increase for 1995/96 alone was 19 percent. During the period that subscription costs rose 57 percent, SIO's recurring collections budget increased just 2.3 percent. Brueggeman singles out Elsevier and Pergamon for particular analysis, finding that "Elsevier titles had a 28 percent increase between 1995 and 1996 and a 32 percent increase between 1992 an 1993. Pergamon titles had a 29 percent price increase between 1995 and 1996 and a 17 percent price increase between 1992 and 1993".

The University of Wisconsin-Madison reports similar effects of serials price increases on its institution.

> Between 1970 and 1990, the cost of journals in chemistry and physics rose by a factor of 12 in current dollars; in psychology, linguistics, and business by a factor of 8 . . . The total campus serials expenditures for 1995 were $4,647,713. One publisher's titles accounted for 17.2% of this figure (almost $800,000), even though this publisher provided only 3% of all serials subscribed to on campus. In the case of the Health Sciences Library, two commercial publishers' titles cost 31% of their budget but represent only 14% of their serial titles. Prices for these journals have been increasing far more than the costs of other Library operations, and double-digit increases are projected for this year.[6]

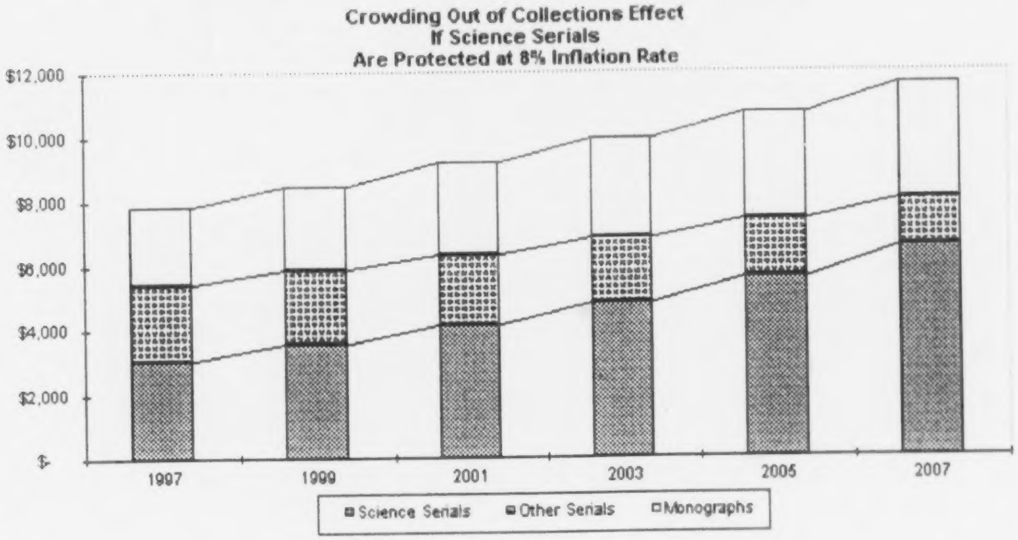
Likewise, at Cornell, Ross Atkinson notes: "While our acquisitions budget was increased this year [1995/96] by a reasonable 4% (the average acquisitions budget increase for the forty largest North American research libraries was 3.7%), the prices of science journals are expected to increase by ca. 18%."[7]

Various authors have demonstrated that not only do the highest cost journals experience the highest rates of inflation, they are also among the most used. Chrzastowski and Olesko[8] found that over a period of eight years, the cost of acquiring the ten most-used Chemistry journals increased 159 percent in comparison to an increase of 137% for the 100 most used journals. During the same period, their usage increased 60 percent in comparison to an increase of 41 percent for the top 100 journals.

Given library budgets that inflate more slowly than the rate of inflation for scholarly journals, there will be a steady decline in the number of titles held in each library. If libraries cancel journals on the basis of use, high-value, high-inflating publishers' titles will be protected, resulting ultimately in a gradual homogenization of collections among libraries. Lesser-used titles, many with low prices and low inflation rates, will be crowded out faster than the general rate of decline in subscriptions held by the library.

The graph below demonstrates a hypothetical scenario. This scenario assumes that the collections budget is inflated by four percent per year and that the expenditures for monographs are inflated at the same four percent rate. However, the average rate of inflation in the cost of scholarly publications is greater. The graph shows that if science journals, because they demonstrate high usage patterns, are canceled more slowly than other titles, and if monograph expenditures are allowed to inflate at the same rate as the overall budget, then science journals will eventually crowd out other journals. In the example, the budget for science journals is allowed to inflate at approximately 8 percent per year (slightly less than one-half the actual inflation rate, but twice the rate of inflation in the total collections budget). Other, lesser-used journals, with lower subscription prices and lower rates of inflation therefore must be canceled more rapidly in order for the collections budget to be balanced. Within a very few years, the high-use/high-price/high-inflation journals could crowd out virtually all other library materials. While no particular library might implement a budget strategy exactly like that depicted in the graph, all libraries tend to retain longest the highest use journals and to cancel first the lesser-used journals. Although the curve may be more gradual, and the time-line longer, the eventual result will be similar to that shown.

Crowding Out Effect
If Science Serials
Are Protected at 8% Inflation Rate



BEST COPY AVAILABLE

Electronic:

There is no evidence that the emergence of electronic journals will change the fundamental economic problems in the cycle of scholarly communication in the short term, at least with respect to commercial publishers. The basic premise of these publishers is that they must both protect their current revenue base and secure guarantees to cover future inflation and increases in content. Thus, publishers frequently structure their initial subscription pricing for digital journals upon the actual cost of paper subscriptions acquired by the institution with which the publisher is negotiating. Often the proposed base subscription rate includes all subscriptions-library, departmental, personal, and other types-identified with the campus, thereby having the effect of greatly increasing the price that the library would have to pay to receive the digital journals. Clearly, publishers are concerned that availability of electronic journals on the campus network will undermine non-library subscriptions to the print versions. After a period of negotiation during which agreements are reached about the institution's existing base cost of print subscriptions, the tougher bargaining begins.

In early 1996, Ann Okerson reported that:

> In general electronic licenses so far have cost on average 1/3 more than print equivalents . . . For full text many publishers also have the expectation that higher price will be asked and should be paid. Publishers are setting surcharges of as much as 35% on electronic journals, and libraries simply do not have the capacity to pay such monies without canceling a corresponding number of the journals of that particular publisher or dipping into other publishers' journals.[9]

Other institutions report that publishers are now agreeing to provide licenses to electronic publications at the same, or marginally increased, price that the institution is paying for print journals. To secure these initially low prices for digital content, the library is asked to consent to such provisos as the following:

1. That there be multi-year (often three) price increase guarantees to compensate for inflation, often at somewhat lower rates than the historical rates for print materials;
2. That there be upward price adjustments for increases in content, often capped at lower rates than typical for print journals;
3. That the publisher be protected against declines in revenue through cancellation;
4. That fair use rights typical for print journals be abrogated for the digital journals.

Although libraries find attractive the ideas both of maintaining a combination of print and electronic subscriptions for a multi-year period without incurring substantial new marginal costs for electronic versions, and of ensuring a "cap" on inflation; neither "feature" of these new licenses will alter the basic economic difficulty in which libraries find themselves: inflation in the price of scholarly information outstrips libraries' ability to pay. In fact, by locking themselves into multi-year agreements that ensure price increases to particular publishers, libraries hasten the rate at which other journals and monographs are crowded out of the market.

Not all scientific publishers have negotiated as described above. For example, both the American Physical Society and the American Mathematical Society offer electronic versions of their journals free to subscribers to the print editions. Clearly, publishers must find revenue streams that will enable them to survive, and the pricing structures for both print and digital journals are the key to those revenue streams. To base a pricing structure for electronic publishing on the costly print model will not be economically viable in the long run (it may, in fact, be unsustainable in the short term as well), as libraries' declining budgets will result inevitably in cancellations to avoid the structural problems associated with double digit inflation, thereby undermining the publishers as well.

The current economic model for scholarly publication cannot be sustained. Continued escalation in the prices for scholarly journals, stagnation in library budgets, and isolation of the creators and consumers of scholarly information (the faculty) from the effects of the economy is the collapse of the system of scholarly communication itself.

Operations Costs in Libraries:

Library operations costs associated with printed scholarly journals include the costs to acquire, process, preserve, and circulate journals. Each library's costs differ based on the organizational structure, degree of centralization and/or decentralization of processes, differentials in salary and benefit scales, effectiveness of automated systems, success at process re-engineering and other factors.

University Libraries and Scholarly Communication reports that "salaries as a percentage of total library expenditures have declined over the past two decades, while 'other operating expenditures' (heavily reflecting computerization) have risen markedly." (p.xxii) While the report infers that the increases in other operating expenditures reflect automation of technical service operations such as acquisition, cataloging, [serials control] and circulation, it simultaneously notes, however, that despite the decline in salaries as a percentage of total library expenses, and the increase in other expenditures, "the number of volumes added per staff member has declined" (p.xxii). Although the decline in acquisitions resulting from inflation certainly affects the ratio of volumes added to staff, it is not possible to discern from ARL statistics the extent to which libraries have programmatically reallocated staff in response to declining receipts and the implementation of automated technical processing and circulation systems. Presumably, greater efficiency in processing and circulation, coupled with declining acquisitions should have resulted in substantial shifts of personnel away from the "back room" of technical processing to provision of direct service to faculty and students. Nevertheless, at least as measured by the ratio of volumes added to staff FTE, it would appear that libraries have not become more efficient overall.

Moreover, University Libraries and Scholarly Communication reports that, on average, library staff increased by a total of 7 percent from 1970 to 1985, and by 6 percent from 1985 to 1991. Thus, the rise in non-salary operations expenses percentage of total operating expenses has not occurred through staff reductions. There has been no systematic study of how the additions in library staff have typically been assigned to various programs. Ironically, the ARL Index ranks research libraries in part on the number of staff they employ; improving productivity and reducing staff accordingly would have the paradoxical effect of reducing a library's ranking vis-a-vis its peers.

The inability to learn from the ARL reports how libraries might be changing their services reflects a serious flaw common to almost all analyses of library costs relating to both collections and operations. Expenditure reports and rankings typically reflect inputs such as volumes acquired, number of serial subscriptions maintained, size of staff, or operational statistics such as the number of circulation transactions, titles cataloged, hours of opening, items borrowed through interlibrary services rather than programmatic outcomes, for example research supported or learning outcomes of students. The problem of defining productivity of knowledge workers was mentioned thirty years ago by Peter Drucker[10], and is further examined by Manual Castells in his recent book, The Rise of the Network Society.[11]

Also lacking in the library literature are large-scale studies of process re-engineering and its effect on the cost structures of particular library operations. Although there has been a great deal of analysis of the costs of materials themselves, and of the scholarly communication system which generates those costs, libraries have thus far been less rigorous in identifying cost centers for processing and other routine library operations. Thus it is not obvious to what extent non-salary investments, for example in automated systems, have actually improved processing productivity or the quality of services rendered by staff; nor is it clear whether or to what degree these investments have moderated the rate of rise of operations costs.

William Massy and Robert Zemsky, discussing the use of information technology to enhance academic productivity in general, remark on its transformational potential, calling it a "modern industrial revolution for the university" which can create economies of scale, deliver broad information access at low marginal cost, and allow for mass customization.[12] The analysis they provide for the academy at large would appear to be even more relevant for libraries, many of whose functions are of a processing nature similar to those in industry, and whose services can also be generalized to a greater degree than is possible for teaching and research.

Massy and Zemsky suggest that although capital investments in technology to enhance productivity will increase the ratio of capital cost to labor cost, they may not actually reduce overall costs. But they offer three major advantages to the shift away from the handicraft mentality resulting from larger capital-labor ratios:

> First, real labor costs tend to rise with economy-wide productivity gains (say two percent per year, on average), whereas technology-based costs tend to decline due to learning-curve effects, scale economies in production, and continued innovation. Increasing technology's share of cost will reduce overall cost growth until the rate differential reduces technology's share to the point where labor again dominates. By this time, however, total cost will be lower than it would have been without the injection of technology. If the real cost of technology were to decline at a 25 percent annual rate, after ten years the alternative scenario would cost about 12 percent less than the baseline. If the rate of decline is only 10 percent, the saving ten years out would have passed 9 percent and still be rising. Given the differential growth rates of labor and technology, one can expect positive long-term returns on investment even when returns are negligible during the first few years.
>
> Second, technology-based solutions also tend to be more scalable than labor-intensive ones. While our model does not address economies of scale, one should expect that additional students could be accommodated at lower cost with technology than with traditional teaching methods.

> Finally, technology provides more flexibility than traditional teaching methods once one moves beyond minor changes that can be instituted by individual professors. The "career" of a workstation may well be less than five years, whereas that of a professor often exceeds 30 years. Workstations don't get tenure, and delegations are less likely to wait on the provost when particular equipment items are "laid off." The "retraining" of IT equipment (for example, reprogramming), while not inexpensive, is easier and more predictable than retraining a tenured profession. Within limits, departments will gain a larger zone of flexibility as the capital-labor ratio grows.

Further, Massy and Zemsky argue: "The benefits of shifting away from handicraft methods, coupled with scale economies and increased flexibility, argue for the adoption of IT even when one cannot demonstrate immediate cost advantages. For example, the ability to break even during the first few years provides strong justification for going ahead with an IT solution, provided the effects on quality are not harmful." Similarly, within the library, use of information technologies, even without generating immediate savings can improve services. For example, online catalogs and automated circulation services provide users with more rapid access to information about the library's holdings, reduce errors in borrowing records, and allow more timely inventory control. Use of online indexing and abstracting services rather than the print versions preserves the scarce time of scholars.

The primary purposes of automating processing operations in libraries have been to reduce the rate of rise of labor costs and to improve timeliness and accuracy of information. Nevertheless, despite the improvements that automation has brought, labor costs to perform library processing operations such as ordering/receiving, cataloging, maintenance of the physical inventory, and certain user services including interlibrary lending and borrowing remain substantial. A transition to electronic publishing of journals would enable libraries to reduce or eliminate many of the costs of these labor-intensive operations, enabling reallocation of the freed-up resources into higher priority services, necessary capital investments in technology, or provision of technology-based information resources. The benefits to end-users would also be significant-less time spent in finding and retrieving physical objects. Ultimately, restructuring of library operations in response to electronic scholarly publishing can, in theory, both improve the quality of services and reduce operations costs. However, to reduce operations costs significantly, libraries will need to define better the desired outcomes of their operations investments, measure those outcomes effectively, and engage in rigorous re-engineering of processes.

There have been several studies which attempt to quantify typical costs of acquiring journals. In a study funded by CLR, Bruce Kingma[13] found the average fixed cost of purchasing a journal subscription to be $62.96. In discussing the economics of JSTOR, Bowen estimates the costs of processing, check-in and binding to be approximately $40.00.[14] In 1996, Berkeley estimated the physical processing costs, including check-in of individual issues, bindery preparation, and binding for print serial subscriptions received and housed in the Main Library, to be as low as $17.47 for a quarterly journal to $113.08 for a weekly journal. Berkeley's figures exclude the costs of ordering and order maintenance under the untested assumption that they will not differ significantly in the case of electronic journals. They also exclude staff benefit costs and overhead and therefore understate the true cost to the university of receiving print subscriptions. Assuming an average annual processing cost of $50.00 per print serial subscription, a research library subscribing to 50,000 titles may incur an operations cost of $2.5 million per year to acquire scholarly journals.

Once the library acquires these journals, it begins to incur the costs of making them available to students, faculty, and other users. In the late 1980's, Michael Cooper reviewed the costs of alternative book storage strategies.[15] He found that circulation costs ranged from a low of $.53 per transaction in a medium sized open-stack research library to a high of $9.36 per transaction from a remote storage facility. Adjusted for inflation of 3 percent per year, these costs would range from approximately $.67-$11.86 per transaction today. Berkeley calculates that an average circulation transaction costs approximately $1.07 and Bowen estimates $1.00. According to ARL Statistics, 1995-96, the mean number of initial circulations per library was 452,428. Using the average circulation transaction cost of $1.00, the average ARL library spent almost $500,000 to circulate materials during FY 1995/96.

Reviewing the costs of acquiring and circulating print journals, it seems fairly obvious that a transition away from acquisition of print and toward electronic journals would reduce annual library operations costs related to providing the university community with the fruits of recent scholarship. Although large recurring expenses in support of historical print collections would continue, they would gradually diminish over time as the aging of the collection reduces the rate of usage. The long-term cost reductions could be substantial in the sciences where currency of information is of utmost importance. Moreover, systematic conversion of high-use print collections to digital form could also generate recurring operations savings. Ultimately, the shift from labor-intensive processing operations to capital investments in electronic content (current journals and retrospective conversion of high-use print collections) could have the kinds of effects envisioned by Massy and Zemsky.

However, caution must be exercised in forecasting these types of potential savings. Despite the potential of long-range savings, they are unlikely to occur to any significant degree in the short term. The pace of transition from print to digital journals is moving slowly, and only those publishers with a strong financial base will be likely to succeed in quickly providing online access. As noted above, and in the section of this paper relating to publishers' cost structures, there is no clearly viable path economically to move to digital publishing. Moreover, libraries will need to maintain print collections, both historical and prospective, into the foreseeable future, requiring that they maintain investments in operations to sustain access to them.

Interlibrary borrowing and lending is a growing cost within research libraries, and its rate of increase promises to escalate as the inflation-generated rate of serials cancellations escalates. According to the ARL, The average annual increase in interlibrary borrowing between 1986 and 1996 was eight percent, and the average annual increase in interlibrary lending was 4.9 percent. Faculty and students borrowed more than twice as many items through interlibrary loan in 1996 as they did in 1986.[16] The University of California Libraries recently reported an annual increase approaching ten percent per year. Interlibrary services are very labor-intensive operations; in 1993, the ARL conducted a cost study which determined the average cost of a borrowing transaction to be $18.62 and that of a lending transaction to be $10.93. The average ARL university library processed 17,804 interlibrary borrowing transactions and 33,397 interlibrary lending transactions during 1995-96, incurring an annual average cost of approximately $700,000. Given the rate of rise of interlibrary resource sharing transactions as well as the rate of rise of labor costs, research libraries are likely to experience increasing interlibrary borrowing/lending costs of at least 10 percent per year.

Capital Costs:

Capital assets in libraries are of three basic types: buildings, collections, and equipment. Expenditures for the most costly of these assets, buildings, are not a part of Library budgets, and therefore are not generally considered by librarians in their discussions of library costs. This paper will not attempt to discuss capital costs for library buildings in any depth except to cite several relevant studies. In the late 1980's Cooper estimated the construction cost per volume housed in an on-campus open stack library to range from $4.33 for compact shelving to $15.84 for traditional open stacks; he calculated the construction cost per volume of a remote regional storage facility to be $2.78. In 1976, Folk[17] estimated the cost of construction to be $4.00 per volume. These costs would be substantially higher today. Bowen uses Cooper's construction costs, adjusted for inflation, and Malcolm Getz' lifecycle estimates, to calculate an annual storage cost of $3.07 per volume. Lemberg's[18] research substantiates Bowen's JSTOR premises regarding the capital cost avoidance possible through digitization of high use materials. He demonstrates that, even considering the capital costs of technology necessary to realize a digital document access system, substantial savings accrue over time, within research libraries *as a system* if documents are stored and delivered electronically rather than in print form. He concludes:

> The results of the various model alternatives for costing the digitized document system and the paper-based document system . . . indicate that very large net present value cost savings can be realized over the assumed model life cycle if a large-scale digitization project is undertaken by academic and public libraries nationwide.

Extrapolating from Bowen's estimate of an annual storage cost of $3.07 per volume, a research library subscribing to 50,000 journal titles per year, each of which constitutes one volume, accrues $153,000 in new storage costs each year. Over ten years the cumulative cost to house the volumes received through the 50,000 subscriptions would exceed $8 million.

The growing dependence on information technologies to deliver scholarly information requires that universities make new investments in capital equipment and allocate recurring operations resources to the maintenance of that equipment and the network infrastructure. Although universities have invested heavily in network technologies, the true costs are still inadequately understood, and it is clear that increasing dependence on digital, rather than print, scholarly information will require that reliable funding models for technology be developed. While capital costs for print libraries entailed buildings and collections, both of whose construction costs fall within known ranges and whose lifecycle is long, capital costs for the digital library are distributed across the campus, and, indeed, the world. However, there is no clear formula to indicate how much initial capital investment in technology might be required to deliver a given number of digital documents to a given size academic community. Moreover, the lifecycle for capital assets relating to delivery of digital library content is typically very short, perhaps as short as five years. Thus capital funding allocations must be made frequently and regularly to ensure continue access. At Berkeley, for example, the Library estimates that annual equipment replacement costs would be approximately $750,000, assuming a five-year lifecycle. But there has never been an explicit capital budget to support that expense, so capital investments in computer equipment, networking, and equipment replacement have been made through redirection of the operating budget. Thus, for the digital library, the library is asked to support, through the operating budget, costs for storage, that, in the print world, are funded from outside of the library's budget. The situation at Berkeley is not unusual, and further work needs to be done to understand more fully the capital cost differentials between the physical plant investments required for print collections and the network investments required to make digital information available to the campus community.

It is possible that if libraries and their parent institutions, universities, could avoid some of the capital and operations costs associated with print-based dissemination of scholarly publications, these resources could be reallocated to capital investments in technology; provision of additional information resources available to the academic community; service improvements within libraries; and restoration of control of the system of scholarly publishing to universities and scholarly societies rather than the commercial sector.

THE ECONOMICS OF ELECTRONIC PUBLISHING: A VIEW FROM THE UNIVERSITY OF CALIFORNIA PRESS

The market realities described in the first portion of this paper are sobering, but the basic outlines have been well known to libraries and scholarly publishers for more than a decade. This section, discusses the realities for nonprofit journal publishers (university presses and scholarly societies) as a way of answering the question, "So why don't publishers just reduce their prices--at least for electronic publications?". Although the focus is on nonprofit presses, the basic economics are equally true for commercial publishers, except that they require profits and have the considerable advantage of greater access to capital to fund innovation.

For all publishers, the largest constraint on their ability to change the price structure for electronic publications radically is the first copy costs-which commonly range from 70 percent to 85 percent of the print price (See Table below for an example of first copy costs for University of California Press journals).

UC Press First Copy Costs
Average, 1994-95
15 February 1997

	Humanitie	Social Sciences	STM	Press-Owned	Contract	Program-wide
Composition/Print	$284,722	$253,005	$50,125	$362,488	$226,127	$588,615
Mailing	$12,696	$15,839	$3,266	$19,792	$12,009	$31,801
Postage	$30,138	$40,970	$5,574	$44,808	$31,874	$76,682
Royalties/Edit. support	$133,602	$797,662	$27,697	$246,796	$712,164	$958,961
Press staff	$254,931	$412,675	$33,290	$401,613	$350,036	$751,649
TOTAL COSTS	$716,089	$1,520,15	$119,952	$1,075,49	$1,332,21	$2,407,708
First copy costs	74.14%	87.94%	71.74%	77.14%	88.22%	83.27%

These first copy costs will remain, whether the format is electronic, paper, or both. Any pricing model must provide sufficient income to cover these costs, in addition to the unique costs associated with publishing in any particular medium. Publishers are not wedded to maintaining print revenues per se but to maintaining enough revenues to cover their first copy and unique-format costs and to covering the costs of the technological shift. In the transition period, when print and electronic editions both must be produced, this will inevitably result in prices that are higher than print-only prices. Whether wholly electronic publications are, in the long run, more economical will depend on the costs of producing uniquely electronic product and on the size of the market. If substantially fewer libraries subscribe to electronic publications than subscribed to their print predecessors, the cost per subscription will inevitably increase in order to cover a larger share of first copy costs.

Electronic Pricing models:

There are a number of models for pricing electronic resources. But all of them ultimately boil down to various ways of obtaining revenue to cover the same set of costs--they all ultimately depend on the same formula of first copy costs plus print costs plus electronic costs.

Let's look at humanities journal x:

Print only

First copy costs	$48,000	80%
Print/paper/bind/mail	$12,000	
Print edition costs	$60,000	
Subscriptions to 1000 libraries @ $60	$60,000	

Print and electronic

First copy costs	$48,000	
Print/paper/bind/mail	$12,000	
Electronic Costs	$12,000	
Total Costs	**$72,000**	
Subscriptions to 1000 libraries @ $72	$72,000	
% increase in total costs		20%

Electronic access provided "free":

Publishers that are providing electronic access "free" with print subscriptions are, in fact, subsidizing the costs of the electronic edition out of the surplus revenues generated by the print publication; the print publication already covers the first copy costs allocated to each subscription. For relatively high-priced scientific journals with high first-copy costs, this can be done without inflating the price too substantially; the uniquely electronic costs are then subsidized by all institutional subscribers and hidden as a percentage of the total cost of publication. Because the basic subscription price is high enough, relatively modest additional increases will also cover the cost of lost individual subscriptions (since individual subscriptions typically cover the run-on costs of producing additional issues but make only a partial contribution to first copy costs). This approach has the added advantage of sidestepping for now the problems of negotiating prices and guarantees with libraries (and the associated overhead costs). However, it does not contribute to developing commonly understood and agreed upon cost recovery models which will cover the costs of electronic scholarly communication in the long run.

Extra charge for electronic access, bundled with paper:

This is essentially the same cost recovery model, but the increase to cover electronic costs is made explicit. This may be especially necessary for journals whose base rate is not so high, so that the markup for electronic costs cannot be covered by a typical inflationary increase. It still has the advantage, for publishers, of spreading the cost over all institutional subscribers and of simplifying licensing negotiations.

Negotiated price by library based on paper subscription base:

This model takes the basic institutional print subscription base and guarantees this revenue for a period of years (typically three). Publishers are willing to guarantee limits to inflationary increases for this period in exchange for the guaranteed income and protection from cancellations to help cover transition costs. Again, this works better with higher priced journals, where the added costs of electronic publishing are a smaller proportion of the total cost.

Separate price and availability for electronic and paper, with an incentive for bundling:

This model-the one basically deployed by SCAN and by Project Muse-offers more flexibility to libraries, since libraries are allowed to cancel print and take only electronic, or to select among the publications offered, although there are discount incentives to encourage maintaining paper and electronic subscriptions (both projects) and/or ordering larger groups of journals (the entire list for Muse; discipline clusters for SCAN). This has the advantage of making the costs of electronic publishing clear. (See the revenues section below for a discussion of the adequacy of this model for supporting humanities publishing in the long run and of the impact of consortia discounts.)

In all these models, the ultimate economic effect in the transition period is the same-costs for libraries go up. Publishers must cover their first copy costs, continue to provide paper editions for individuals, many libraries, and international markets, and to generate revenue to cover the infrastructure and overhead costs of electronic innovation. For nonprofit publishers, at least, these costs must all be supported by the revenues from current journal subscriptions.

Electronic costs:

It is likely, in the long run, that eliminating print editions entirely will reduce costs somewhat for some kinds of journals. However, for journals which are trying fully to exploit the new capabilities offered by electronic technologies, it seems likely that the additional costs of generating links, specialized formats, etc. will continue to cost as much, or nearly as much, as the cost of printing and binding. (See The Astrophysical Journal at http://www.journals.uchicago.edu/ApJ/, Earth Interactions at http://earth.agu.org/ei/, or any humanities journal with lots of multimedia). But even for simpler humanities journals, the experience at the University of California Press raises questions about the assumption that ongoing electronic costs will be substantially lower.

Covering costs of development:

The University of California Press' original economic model assumed that the development costs were largely one-time expenses, that there was a single learning curve and set of expertise to master, after which electronic publishing would be largely routinized; additional expenses would be easily absorbed by the margin generated by the savings in the paper edition. On the basis of the past three years, it seems apparent that this was a flawed assumption. UC Press dedicated 3,500 staff hours on the SCAN project in 1994 (gopher site development); 4,100 hours in 1995 (WWW site development); and 3,700 hours in 1996 (largely on WWW development and on laying the groundwork for SGML implementation). It is apparent from ongoing trends in technological innovation that Internet technology and expectations for electronic publishing will continue to evolve very rapidly for at least the next twenty years. The Press' "bad luck" in initially developing for an outmoded platform (gopher) is an inevitable occurrence over the long-term for electronic publishing projects. As a result, it seems foolhardy to assume that there will be substantially less investment necessary for technical research, experimentation, and site redesign and revision in the future. Any viable economic model for the University of California Press must thus assume one or two technical FTE positions as part of ongoing overhead (please note, this does not include file server maintenance and enhancement, since the costs of file-service are presently borne by University of California/Berkeley Library for the SCAN project).

In addition, the SCAN project has experienced ongoing instability in technical staff-at the Library and at the Press. Being located in a region with such a strong high technology industry has actually proven to be a disadvantage, since current and potential employees can make so much more money at other jobs. This results in long staff vacancies and with repeated training on the specifics of the project. It's another way in which there is not one but rather a continual series of learning curves.

There is a third implication to this vision of a continually changing future. Combined with the Press' commitment to long-term responsibility for viable electronic access and to archiving, continually changing platforms and functionality demand implementation of a coding system which is totally portable and highly functional. As a result, the commitment to SGML seems more and more wise as time goes on. This commitment leads the Press to reject image-based solutions like Acrobat which would be less work and which would be faster to implement but which do not have long-term migration paths. Having once lived through the painful process of having to completely re-code each individual file, the Press does not want to face the same problem with a much larger set of files in the future. The necessity and the difficulty of repeated conversions of legacy text is currently sadly underestimated by many publishers and librarians. Scaleability-an important and underrated issue in any case-becomes even more vital in a scenario in which larger and larger amounts of material must be converted each time the technological environment evolves.

In addition, electronic publishing is adding new duties (and requiring new resources) within the Press, without removing present duties. For example, the Press has added .5FTE in the journals production staff (a 25 percent increase) to handle liaison with suppliers, scanning and archiving of all images being published, archiving of electronic files, and routine file conversion duties. This position will clearly grow into a full-time position as all the journals are mounted online; only the slowness of the online implementation permits the luxury of this low staffing level. The seven people working on Project Muse or the seven people working on The Astrophysical Journal Electronic Edition confirm this assumption. In addition, clearing electronic rights for images in already-published books and journals and maintaining an ongoing rights database creates an ongoing staff responsibility, since many rights holders are requiring renewal of rights and payments every five to ten years. This is a wholly new function which must be incorporated into ongoing job functions and overhead. The need for technical customer support is still essentially unknown but surely represents some portion of an FTE.

Marketing is another area requiring addition of new expertise and staff. Successfully selling electronic product requires a series of changes within the publishing organization. The marketing necessary to launch a new print journal successfully or to sell a book is expensive and time-consuming, but the approaches and tasks are familiar and can be performed by existing marketing staff as part of their existing marketing jobs. In contrast, successfully establishing a customer base of licensed libraries for electronic product requires new skills and abilities, a substantial staff commitment, a higher level of staff expertise and authority, and substantial involvement from the licensing library. Marketing electronic services requires all the brochures and ads that print publications do. In addition, it requires substantial publicity efforts, a travel schedule to perform demonstrations at a wide range of library and end-user meetings, participation in appropriate listservs, and at least one staff member who has the requisite knowledge and authority and who can dedicate a large portion of their time to outreach, negotiations, and liaison with potential and actual license customers and subscription agents. There are also demands for ongoing customer relations work, including the provision of quarterly or annual use reporting. The Press has found it very difficult to fit those new functions into its traditional marketing and distribution job descriptions and workloads. As the Press moves more seriously into electronic publication of frontlist books, it will surely need to hire a new person to market online books; it will not be possible to integrate these functions into the already busy jobs of books marketing professionals with their focus on current season bookstore sales.

In short, the Press anticipates a permanent addition of at least three or four full-time staff to the overhead of the publishing operation. For now, some of these positions are covered by the Mellon Foundation grant, and some of them have been deferred (to the detriment of the project), but in the long run the electronic publishing model must absorb these additional $200,000 in annual costs.

Finally, the Press and the Library have just begun to step up to the costs of long-term archiving (including periodic refreshing of technology and the requisite reconversion of files-another argument for structured standardized coding of text).

Income for electronic product:

Unfortunately, in a period when electronic publishing generates additional costs which must be funded, there are several trends apparent in the emerging purchase patterns of electronic products which limit the income available to support publication costs and which create further pressures on publishers to increase prices.

Slowness to adopt:

University presses which are attempting to sell electronic product directly (as opposed to bundling it automatically in the paper price, and offering "free" access to the electronic product) are finding that sales to universities are progressing more slowly than projected. Project Muse sales, for example, are at 378 after 2 years; sales to MIT's electronic-only journals hover at around 100; in no case are there more than fifty library subscriptions. There are under 25 subscriptions to the online edition of The Cigarette Papers at the University of California/San Francisco Library's Brown and Williamson Tobacco site after nine months (http://www.library.ucsf.edu/tobacco/cigpapers/). Sales to SCAN are a handful (although access has been restricted for less than one month at the time this paper is written). Even for publications for which no additional charge is being made, library adoptions are still slow in coming-The Astrophysical Journal Electronic Edition, for example, has 130 libraries licensed to date. There are, of course, good reasons for this slowness; libraries face the same difficulties in building infrastructure, funding, and staff expertise that publishers do. But the low sales nevertheless make funding the transition more difficult, because publishers can't count on sales income from the electronic product to help to cover the costs of electronic publication. The growth curves to which publishers are accustomed from launching paper journals (even in this age of low library adoptions) are too optimistic when applied to electronic publications. This has real consequences for funding electronic innovation.

New discount structures:

In addition, the emerging business practices and discount expectations lessen the income per subscribing institution (at the same time as the efforts necessary to obtain that subscription are intensified). The expectations of consortia for deep discounting (both for number of consortia members and for adopting a bundle of publications) can go as high as 40 percent for academic institutions, with non-traditional markets receiving even deeper discounts. If one assumes that the 70-85 percent of the list price represents the first copy costs, a 40 percent discount means that these subscriptions are no longer carrying their full share of the first copy costs. This can't be a long-term pricing strategy.

In addition, there are often other consortial demands (for example, demands that inflationary increases not exceed a certain percentage for several years, or that access be provided to high schools free of charge) which further lessen the ability of publishers to fund electronic innovation out of electronic product sales. Again, it is easy to empathize with these library positions and to understand why they are evolving. But these efforts by libraries to control costs actually have an inflationary pressure on overall prices, since the base price must increase to make up the losses.

Loss of subscriptions:

In addition, publishers are worried about losing subscriptions. Some losses will surely happen:

another major wave (or waves) of cancellations as libraries try to cope with the ongoing costs of paper and electronic subscriptions from the major commercial science publishers; and the loss of any duplicate subscriptions still remaining on campuses. In addition, publishers are haunted by the potential for substantial shrinkage of individual subscriptions/society memberships as more and more scholars have "free" access from their campuses, though loss of individual subscriptions is less sure than library cancellations (by December 1996, almost 60 percent of SCAN uses were coming from domestic non-.edu addresses as more and more people obtain access from home workstations; it is possible that individuals will pay for the convenience of non-campus access, just as they now do for non-library print access.) Nevertheless, because individual subscriptions play an increasingly important role in financing many journals (especially journals launched within the past ten years, when library support has been so eroded), widespread cancellation would have a substantial impact which would force journal prices higher.

Possible increases in sales:

There are two possible new revenue sources that may somewhat balance the losses in income described above, although both are highly speculative at this point. First, publishers may obtain new institutional markets and wider distribution as consortia bring institutions like junior colleges and high schools to scholarly publications. Project Muse has begun to see this trend. It is not clear, however, that these will be long-term subscribing customers. Given the present nature of scholarship, many of these new subscribers may conclude that any amount of money is too much to pay after two or three years of very low use statistics, especially when by-article access on-demand becomes widely available. There will be a substantial market for scholarship at junior college, high school, and public libraries only when the possibility of wider audiences through the Internet fundamentally changes the ways in which scholars write and present their work--a change that will surely take many years to materialize. Other publishers are more optimistic about this potential source of income.

Second, there may be a substantial revenue stream in sale of individual chapters and articles to scholars whose institutions do not have access, who do not have an institutional base, or who are willing to pay a few dollars for the convenience of immediate access at their workstations (people who are now presumably asking their research assistants to make photocopies in the stacks). And there may be substantial sales among the general public. This new product may represent substantial income which could relieve some of the pressure on journal finances, if the process can be entirely automated (at $6 or $7 per article, there is no room for the cost of an employee ever touching the transaction). There will need to be substantial traffic here, as it takes seven or eight article sales to cover the first copy costs of one typical humanities subscription.

Of course, the ability to purchase single chapters or articles will also diminish subscription revenues, as some libraries choose to fill user needs on demand and to cancel their present subscriptions. It is too soon to tell what the mix of new audiences and subscription cancellations will be, and whether the revenue stream from new sources will replace that from canceled subscriptions.

Aggregators:

So far, the models we have examined have all assumed that the publisher is providing access to electronic resources. Publishers could, of course, avoid many of these costs by providing electronic files to aggregators and leaving marketing, file-service, file conversion, and archiving to outside suppliers who would provide a single point of purchase for libraries and individuals. This scheme offers a number of advantages from a library point of view. The instant connection between search engine and ordering ability which the larger services like Uncover and OCLC offer may potentially bring more end-users.

But from a publishing point of view, there are two very large disadvantages. The first is strategic. In an electronic world, one of the major values which publishers have to offer is the branding value of our imprints as symbols of excellence resulting from peer review and gatekeeping-functions which will be ever more valuable in the time-starved world of the Internet. This brand identity is inevitably diluted in an aggregated world, especially if the aggregator is handling marketing and distribution.

Second, and more relevant to the discussion at hand, it is hard to see how the royalties most typically offered by aggregators (for institutional licenses or for on-demand use) can begin to replace the revenue lost from direct subscriptions. A 30-40 percent royalty does not cover first copy costs of 80 percent. Only by retaining the entire fee can publishers hope to generate enough revenue for on-demand sales to make a sufficient contribution to the costs of publication. A wide-scale move to aggregation would have the effect of making the first copy costs for the few remaining subscriptions very large indeed, in addition to reducing the perceived value of what we sell (yes, it is possible for a humanities quarterly to cost $1200 annually!) .

The University of California Press and most other nonprofit scholarly publishers would like nothing better than to price electronic products substantially lower than print. However, the low margins under which they operate, the demands of users that print continue to be provided, the high first copy costs typical of scholarly publishing, the need to fund the development of electronic product, and the expenses of producing full-featured electronic publications all mitigate against low prices, at least during the transition period.

CONCLUSION:

The university press and the library face economic pressures that neither can address alone. In the face of continuously escalating prices and relatively flat budgets, libraries will continue to reduce acquisition rates to balance the collections budget, and these reductions will adversely affect the revenues to university presses. In addition, the pressure from the sciences, technology, medicine and business to retain high-cost, high-use journal subscriptions will tend to crowd out lesser used scholarly journals, many of which are published by university presses. The need to maintain large physical plants and control large print inventories will continue to mitigate against libraries' employing the kinds of radical, cost-reducing changes in operations that could free up resources for investments in technology. The trends noted in University Libraries and Scholarly Communication, and in Hawkins' paper will result in a catastrophic decline in the system of scholarly communication unless there is a fundamental shift in the way in which its processes, products, and costs are analyzed. Each of the two partners, the library and the press, serves as an inadequate unit of analysis for the system of scholarly communication as a whole.

Sandra Braman's description of the three stages in the conceptualization of the information society[19] provides a useful context in which to view today's problems of press and library within the system of scholarly communication. She characterizes the three stages of conceptualization as follows. In the first stage, although the economy is seen to be operating normally, it is recognized as an information economy because industries in that sector are of greater importance than in the past. The second stage is characterized by commodification of forms of information never before commodified. In this stage, political controversy about information's value as a public good vs its market value as a commodity is highlighted.

In the third stage conceptualization, a more sophisticated understanding of the flow of information replaces the market as the primary feature of the information economy. This stage represents a paradigm shift in which the information economy is seen to operate in a *qualitatively* different manner than in the two previous conceptualizations. According to Braman: "key insights of this perspective include identification of a new unit of analysis, the project, involving multiple interdependent organizations, as more useful than either the industry or the firm for analytical purposes."(p. 112) She further describes the third stage conceptualization of the information economy as including a production chain-or "harmonized production flows" including information creation, processing, storage, transportation, distribution, destruction, seeking, and use-in short, all of the stages of the system of scholarly communication from author to user, including the library. In the third stage, networked information economy, economic viability stems not from maximizing profit or economic stability within each component of the system, but rather through building long term relationships and a stable *system* or *flow* of information.

Michael Hammer makes a similar point with respect to industrial or business reengineering, but applicable to libraries and presses as well:

> The usual methods for boosting performance-process rationalization and automation-haven't yielded the dramatic improvements companies need. In particular, heavy investments in information technology have delivered disappointing results-largely because companies tend to use technology to mechanize old ways of doing business. They leave the existing processes intact and use computers simply to speed them up . . . Instead of embedding outdated processes in silicon and software, we should obliterate them and start over. We should "reengineer" our businesses: use the power of modern information technology to radically redesign our business processes in order to achieve dramatic improvements in their performance.[20]

Both Braman and Hammer emphasize the disquieting qualities that characterize this kind of paradigm shift implied by the third stage of conceptualization of the information economy and by successful reengineering. According to Hammer,

> Reengineering cannot be planned meticulously and accomplished in small and cautious steps. It's an all-or-nothing proposition with an uncertain result . . . At the heart of reengineering is the notion of discontinuous thinking-of recognizing and breaking away from the outdated rules and fundamental assumptions that underlie operations. Unless we change these rules, we are merely rearranging the deck chairs on the Titanic. We cannot achieve breakthroughs in performance by cutting fat or automating existing processes. Rather, we must challenge old assumptions and shed the old rules that made the business under perform in the first place . . . Reengineering requires looking at the fundamental processes of the business from a cross-functional perspective.

Manuel Castells takes a different approach, suggesting that technology-driven productivity increases in the *informational economy* have not thus far been evident. His thesis is that while technology-driven productivity increases were steady in the industrial sector between 1950 and 1973, since 1973, despite the intensive investment in technology, productivity-- particularly in the service sector has stagnated. He suggest three factors which appear to be relevant to the library/press sector, as well as to the service sectors of the economy in general. These factors include the following.

> 1. **Diffusion**: before technological innovation can improve productivity markedly, it must have permeated the whole economy, including business, culture, and institutions.
>
> 2. **Measuring productivity**: Service industries traditionally find it difficult to calculate productivity statistically; thus the lack of observable productivity enhancements may in part be a symptom of the absence of relevant measures.
>
> 3. **The changing informational economy**: Productivity cannot easily be measured because of the broad scope of its transformation under the impact of information technology and related organizational change.

If Castells and Braman are correct, then libraries and presses, alone or together, cannot implement technological solutions that can transform the processes, productivity and economics of scholarly publishing.

The Mellon projects have been useful in introducing two players in the information flow to the problems of the other, and in forging collaborative relationships to aid in sustaining the system of scholarly communication. These cooperative projects between university libraries and presses are useful in helping participants to begin to understand the system of scholarly publishing as an information flow rather than as separate operational processes. But they are limited in effectiveness because outside of the parameters of the projects, the partners must still maintain their separate identities and economic bases.

In a fuller exploration of the potential of transforming the flow of scholarly information, there would be a more integrated economic model including the creators of the information as well as the publisher, the library, the university administration, and the consumers. In this system, costs and subsidies of the entire process of scholarly communication would be better understood, and resources made more flexibly available to support it. For example, it might be possible to view operational and capital savings to libraries resulting from a transition to electronic publication as resources ultimately available to sustain the publication chain, or consumers could be asked to pay some of all of the costs of creating, storing, archiving, and delivering scholarly information. A critical flaw in the current system is the existence of a part of the gift economy, in the form of the library, within a monetary economy for commercial publishers. Because the consumers of the information within the university do not pay for it, they and the campus administration see the library as a "problem" when it cannot provide the information needed within the budget allotted.

A key problem in securing the future of scholarly communication is that both presses and libraries are undercapitalized. Although libraries incur huge capital costs over time-in both inventory and facilities, they are not free individually nor as parts of the *system* of scholarly communication to reallocate present or future capital expenditures to investments in new modes of publication. However, such reallocation, if it occurs at all, will take place very slowly because the transition to digital publication will also be slow. It is possible that a more rapid transition to electronic publishing would reduce libraries' recurring operations costs, thereby enabling them to invest greater resources in information itself. But a more rapid transition is feasible for presses only if there is a rise in demand for digital publications from libraries and from end users or a substantial increase in subsidies from their parent universities. Presses can offer electronic publications, but they cannot change the demand patterns of their customers, libraries, nor the usage patterns of the end consumers in order to hasten a transition from print to electronic dissemination. As long as a substantial portion of their market demands print (or fails to purchase electronic product), presses will be forced to incur the resulting expenses, which, in being passed on to libraries as costs that inflate more rapidly than budgets, will reduce the purchases of scholarly publications..

Ironically, in the present environment, universities tend to take budgetary actions that worsen the economics of scholarly communication as experienced by both libraries and presses. University administrators increasingly interpret any subsidy of university presses as a failure of the press itself as a business; as university subsidies are withdrawn, presses must increase prices, which reduces demand, and exacerbates the worsening fiscal situation for the presses. But in the networked economy where everyone can be an author and publisher, the value added by presses (for example gatekeeping, editorial enhancement, distribution) may be more important than ever in helping consumers select relevant, high quality information. At the same time, university administrators see the library as a "black hole" whose costs steadily rise faster than general inflation. Since library materials budgets grow more slowly than inflation in the costs of scholarly publications, the inevitable result is reduced purchasing of scholarly publications of all types, but particularly of university press materials which in general are of lesser commercial value in the commodity market. Unless the system as a whole changes, both university presses and university libraries will continue to decline, but at an increasing rate.

Although it is not possible to envision with certainty exactly how a successful transition from the present system to a more sustainable system might occur, one plausible scenario would be for universities themselves to invest capital resources more heavily in university-based information flows and new forms of scholarly publication as well as for them to place increased market pressures on the commercial sector. If universities were to make strategic capital and staffing investments in university presses during the short term, the presses could be more likely to make a successful and rapid transition to electronic publication. At the same time, intensive university efforts (i.e. investments) to recover scientific, technical, medical, and business publishing from the private sector could be made to reduce the crowding out of university press publications by for profit publishers. These efforts to recover scholarly publishing could be accompanied by libraries' placing strong market pressures on commercial publishers through cancellation of journals whose prices rise faster than the average rates for scholarly journals in general. The investments in these two areas-converting publication processes to electronic form, and returning commercial scholarly publishing to the university-could be recovered over time through reductions in capital investments in library buildings. Ultimately, the university itself would encompass most of the information flow in scholarly communication through its networked capability. That information having commodity value outside of the academy could be sold in the marketplace, and the revenues used as a subsidy to the system itself.

Another way of accomplishing a harmonization of the scholarly information economy was suggested by Hawkins: the independent non-profit corporation model[21] in which universities and colleges would invest together in a new organization which would serve as a broker, negotiator, service provider, and focus for philanthropy. It would leverage individual resources by creating a common investment pool.

However the solution to the problem of the economic crisis in scholarly communication is approached, there must be a fundamental change in how the process as a whole is conceived, and how intellectual property rights of both authors and universities are managed. Such a change cannot be made unilaterally by university libraries and presses, but will require the strategic involvement and commitment of university administrators and faculty within the university and among universities. Patricia Battin, envisioning an integrated scholarly information flow said almost ten years ago:

> Commitment to new cooperative interinstitutional mechanisms for sharing infrastructure costs -- such as networks, print collections, and database development and access -- in the recognition that continuing to view information technologies and services as a bargaining chip in the competition for students and faculty is, in the end, a counterproductive strategy for higher education. If the scholarly world is to maintain control of and access to its knowledge, both new and old, new cooperative ventures must be organized for the management of knowledge itself, rather than the ownership of formats.[22]

FOONOTES:

[1] Battin, Patricia. "Libraries, Computers and Scholarship." Wilson Library Bulletin, April 1982, 580-581.

[2] University Libraries and Scholarly Communication: a Study Prepared for the Andrew W. Mellon Foundation, Washington, DC, Published by the Association of Research Libraries for the Andrew W. Mellon Foundation, 1992.

[3] Hawkins, Brian. "The Unsustainability of the traditional library and the threat to higher education," paper presented at the Stanford Forum for Higher Education Futures, The Aspen Institute, Aspen Meadows, CO, October 18, 1996.

[4] "Price Comparison of STM Journals 1996/1997" Harrassowitz News: Press Release (November 1996) URL: http://www.harrassowitz.de/news/9611pr01.html.

[5] Brueggeman, Peter. "Journal Subscription Price Increases," URL: HTTP://scilib.ucsd.edu/sio/guide/five-yr.html, 2/9/97.

[6] University of Wisconsin-Madison. "University Library Committee Proposed Resolutions on Faculty Concerns on Copyright and the Role of Libraries (as adopted and amended May 6, 1996). Faculty Document 1214a, 6 May 1996. URL: http://www.library.wisc.edu/libraries/issues/ulc-ipr.htm

[7] Atkinson, Ross. "Summary of the 16 November 1995 Presentation to the Science Faculty.

URL: http://scilib.ucsd.edu/sio/guide/five-yr4.html

[8] Chrzastowski, Tina E. And Brian M. Olesko, "Chemistry Journal Use and Cost: Results of a Longitudinal Study," Library Resources and Technical Services, v.41, no.2, 1997 p. 101-111.

[9] Okerson, Ann. "A Librarian's View of Some Economic Issues in Electronic Scientific Publishing," Paper presented at the UNESCO Invitational Meeting on the Future of Scientific Information. Paris, February 1996. (URL: http://www.library.yale.edu/~okerson/unesco.html)

[10] See, for example, Drucker, Peter. The age of discontinuity; Guidelines to our changing society. New York, Harper & Row, 1969.

[11] Castells, Manuel. The Rise of the network society. Cambridge, MA, Blackwell, 1996. See especially chapter 2, "The informational economy and the process of globalization", which discusses "the productivity enigma".

[12] Massy, William F. And Robert Zemsky. "Using information technology to enhance academic productivity". URL: http://www.educom.edu/program.nlii/keydocs/massy.html

[13] Kingma, Bruce R. "The Economics of Access versus Ownership: The Costs and Benefits of Access to Scholarly Articles via Interlibrary Loan and Journal Subscriptions". URL: http://www-clr.stanford.edu/clr/econ/kingma.html (last updated January 27, 1997).

[14] Bowen, William G. "JSTOR and the Economics of Scholarly Communication 1" URL: http://www-clr.stanford.edu/clr/econ/jstor.html

[15] Cooper, Michael. "A Cost Comparison of Alternative Book Storage Strategies," Library Quarterly, v.59, no.3, 1989 p. 239-260.

[16] ARL Statistics, 1995-96. Washington, DC, Association of Research Libraries, 1997. p. 11.

[17] Folk, Hugh. "The Impact of Computers on Book and Journal Publication", in Proceedings of the 1976 Clinic of Library Applications of Data Processing: The Economics of Library Automation, ed. By J.L. Divilbiss. Urbana, IL, University of Illinois Graduate School of Library Science, 1977. p. 72-82.

[18] Lemberg, William Richard. A Life-Cycle Cost Analysis for the Creation, Storage and Dissemination of a Digitized Document Collection. Dissertation (Ph.D.), School of Library and Information Studies, Berkeley, 1995. Also available online as: URL: http://SIMS.berkeley.edu/research/publications/DigitDoc.pdf

[19] Braman, Sandra. "Alternative Conceptualization of the Information Economy," in Advances in Librarianship, v. 19, 1995, p. 99-116.

[20] Hammer, Michael. "Reengineering Work: Don't Automate, Obliterate." Harvard Business Review, July-August 1990, p. 104-112.

[21] Hawkins, Brian. "Creating the Library of the Future: Incrementalism Won't Get Us There! Serials Librarian, v. 24, no.3/4, 1994. P. 17-47.

[22] Battin, Patricia. "New Ways of Thinking About Financing Information Services," in Organizing and Managing Resources on Campus, ed. Brian L. Hawkins (McKinney, Texas: Academic Computing Publications, 1989), 382.

……

For additional information about the conference, or The Andrew W. Mellon Foundation's scholarly communication initiatives, please contact Richard Ekman. For additional information about ARL or this web site contact Patricia Brennan, ARL Program Officer at (202) 296-2296.

Return to Office of Scholarly Communication Home Page

ARL Home ARL Scholarly Communication and Technology Home Page

© Association of Research Libraries, Washington, DC

Web Design by Angelo F. Cruz
Maintained by ARL Web Administrator
Last Modified: August 29, 1997

Scholarly Communication and Technology

Conference Organized by The Andrew W. Mellon Foundation

at Emory University
April 24-25, 1997

Copyright © of the papers on this site are held by the individual authors or The Andrew W. Mellon Foundation. Permission is granted to reproduce and distribute copies of these works for nonprofit educational or library purposes, provided that the author, source, and copyright notice are included on each copy. For commercial use, please contact Richard Ekman at the The Andrew W. Mellon Foundation.

Session #1 Economics of Electronic Publishing: Cost Issues

Economics of Electronic Publishing: Cost Issues - Comments on Session One Presentations

Robert Shirrell
Journals Manager
The University of Chicago Press

• •

I have a few brief comments on the very interesting and stimulating talks we've heard by Janet Fisher, Malcolm Getz, and Bill Regier. I'll focus on their presentations of publisher costs, and I'll add a few words about the electronic publishing efforts we have undertaken at the University of Chicago Press and contrast the model we have adopted with the ones that have been mentioned earlier.

Janet Fisher, from the MIT Press, gave us costs related both to the electronic journals that they are publishing and to two of MIT's print journals. In Table One I've reworked the numbers and computed "first-copy" costs on a per-page basis. What I mean by "first-copy cost" is simply the cost for editing, typesetting, and producing materials that can subsequently be duplicated and distributed to several hundred or several thousand subscribers. The total first-copy costs for electronic journals at MIT Press range from approximately $15 to $56 per page, and the total first-copy costs for the print journals are $22 and $24 per page. In computing these costs, I did not include what Janet labeled as the "G&A" costs, the general and administrative costs, but I did include the portion of the cost of the Digital Lab that is related to first-copy production.

There are several things here that I think are important and worth a comment or two. First, the Digital Lab cost, the cost of preparing an electronic edition after editing and typesetting, is a significant portion of the total . Although the percentage varies between 13% and 62% (as indicated in Table One), the cost is close to 50% of the total first-copy costs of publishing these particular electronic journals.

This breakdown raises the questions, Why are these costs so high? and Will they decline over time? I think the expense reflects the fact that there are hand-crafted aspects of electronic production, which are expensive, and there are substantial hardware costs that need to be allocated among a relatively small number of publications and a small number of pages. As for the future, the per-page costs at the Lab can be expected to go down as pages increase and new processing techniques are developed, but even if they do go down to 40%, the totals for the digital production are going to be a significant portion of the publisher's total cost. This is important.

Another point about these costs. Note that the total first-copy costs of the electronic journals average $40-$43 per page, and those for the print journals average about $23 per page, roughly a $20 difference in the costs. For a 200 page issue, that would amount to about $4,000. That is, it is $4,000 more expensive to produce materials for reproduction and distribution of 200 pages in electronic form than it is to produce materials for reproduction and distribution of 200 pages in hardcopy form.

If $4,000 will pay for printing and distribution of a 200-page issue to 500 subscribers, which is a reasonable estimate, then MIT can produce a print edition less expensively than an electronic edition when the distribution is under 500. That is an important conclusion: At this point, for the MIT Press, it's cheaper to produce journals in paper than to do them electronically, if the circulation is small. That may evolve over time, but right now, it's still cheaper to be in print until circulation rises to at least 500, because for small-circulation totals the additional costs of electronic processing are not offset by sufficiently large reductions in printing and distribution costs.

Now let me turn to the presentation by Malcolm Getz. Malcolm presented some numbers from the American Economic Association (AEA), and the numbers in Table Two are approximately the same as the ones he presented. I have also presented numbers from the University of Chicago Press for 37 of our titles. That is not the total of our serial publications - we publish 54 in all. It excludes *The Astrophysical Journal*, our largest single title, and a number of journals that we publish in cooperation with other not-for-profit organizations. The journals that are included are principally titles in the humanities and social sciences, with some in medicine and biology.

The breakdown of costs for the Press and for the AEA is quite similar. Editorial costs are 36% for AEA and 32% for the Press. Typesetting is 13% for AEA and 10% at the Press, though it varies substantially by journal. Distribution costs are similar. Overall, these numbers are very close, and they are, it seems to me, reasonable numbers industry-wide.

It is possible to provide a more detailed break-down of the numbers for the Press, and in Table Three I have broken down the 32% that is related to editorial into the portion that is related to the peer review of manuscripts, which is 22% of the total, and the portion that is related to manuscript editing, which is 10% of the total. Because of the manner in which some of the Press's costs are recorded, the number I have shown for manuscript editing may be somewhat higher, but the breakdown between peer review and manuscript editing is a reasonably accurate division of costs in traditional journal publishing. I think this revised breakdown of costs provides an interesting context for reviewing the way in which costs evolve in an electronic publishing environment, and I would like to turn now to make a few remarks about the possibilities for cost restructuring and cost reduction.

The electronic publishing model we have been discussing this morning is structured so that, basically, electronic costs are add-on costs - you do everything you do in print, and then you do some more. I have outlined the process in Table Four. The process includes the traditional functions of peer review, manuscript editing, typesetting, printing and mailing, and adds new functions and new costs for the derivation of electronic materials from the typesetting process and for the management of electronic services.

In this model, for the vast majority of journals, as long as we continue to produce both print and electronic editions, the total cost is not going to decrease. The reason is that, even if a significant portion of the subscribers convert from paper to electronic editions, the additional costs for electronic processing are not offset by reductions in the printing and distribution costs. As we all know, the marginal cost of printing and mailing is small, much smaller than the average cost, and the additional costs for electronic processing are substantial. The consequence is that, in this model, electronic costs turn out to be added costs, costs in addition to the total that would exist if only a print edition were being produced.

This is exactly what we heard from Bill Regier. He reported that for Project Muse, the electronic publishing venture of the Johns Hopkins University Press, the total costs for both print and electronic editions were about 130% of the print-only costs. This is a significant increase, and I believe it is representative of efforts that are based on deriving electronic materials from typesetting files, as a separate stage of production, undertaken subsequent to the typesetting process.

I would now like to discuss another approach to electronic publishing, another way to obtain electronic materials and to do electronic dissemination. This process is quite different from the one I have just described, with different cost structures and different total costs. The process is outlined in Table Five. In this process, data are converted to SGML form in the earliest stages of editing. Then the SGML database is used to derive both the typeset output for hardcopy printing and the electronic materials for electronic dissemination.

This process generates costs quite different than those for the model we looked at before. The costs are summarized in Table Six. Most important, there is a substantial increase in the cost at the beginning of the process, in the conversion of data to SGML form and the editing of it in that format. SGML editing is not easy and it is not cheap. However, because manuscripts are extensively marked up and formatted in this process, a typeset version can be derived from the SGML database inexpensively, and of course, the electronic files for distribution in electronic form are also straightforward and inexpensive to derive. Overall, the additional costs for conversion and editing are being offset in large part by reductions in typesetting costs.

This is the process that we have undertaken with *The Astrophysical Journal* at the University of Chicago Press and are now implementing for other publications. *The Astrophysical Journal*, sponsored by the American Astronomical Society, is the world's leading publication in astronomy, issuing some 25,000 pages each year, in both print and online editions. The conclusions we have reached in our efforts for that journal are that a reduction in the typesetting costs can offset other additional costs, and that this method of producing the journal is less expensive than any alternative way of generating the electronic materials that we want to obtain for the online edition.

These general conclusions are probably applicable to most scientific and technical journals, as this method - based on processing in SGML form - results in substantial reductions in the cost of typesetting tabular and mathematical matter. For those publications, we will be able to produce electronic editions for at most 10% more than the cost of producing print editions alone. In some cases it may be possible to produce electronic versions, in addition to the print versions, at no additional total cost.

Let me add one other point. Because we are converting manuscripts to SGML immediately and editing in SGML, we can obtain materials for electronic distribution much faster than in the traditional print model. Later this year we will publish papers in the online edition of *The Astrophysical Journal Letters* 14 days after after acceptance by the editor. That is possible because we will obtain the electronic version immediately from our SGML database and not derive it by post-processing of typesetting files.

In sum, with this process, in certain circumstances, we will be able to publish complex scientific material in a sophisticated electronic version both less expensively and more rapidly than by employing alternative means. This sort of processing is an important alternative approach to electronic publishing.

Table One

MIT Press First-copy Cost per Page

Electronic Journals

	JFLP	SNDE	CJTCS	JCN
MS Editing		7.25	4.57	
Composition		18.20	8.48	
Subtotal	7.87	25.44	13.05	49.00
Lab	7.68	18.42	21.31	7.00
Total	15.55	43.86	34.35	56.00
Lab %	49%	42%	62%	13%

Print Journals

	NC	COSY
MS Editing	6.46	6.93
Composition	16.04	17.57
Subtotal	22.50	24.50
Lab		
Total	22.50	24.50

Table Two

Cost Breakdown by Percentage for

AEA (3 journals) and University of Chicago Press (37 journals)

	AEA	Press
Editorial	36%	32%
Typeset	13%	10% (to 18%)
Print and Mail	23%	24%
Other	27%	34%

Table Three

Cost Breakdown by Percentage for

University of Chicago Press (37 journals)

Editorial	
Peer Review	22%
MS Edit	10%
Typeset	10% (to 18%)
Print and Mail	24%
Other	34%

Table Four

Cost Breakdown for Electronic Publishing Model One

Editorial		
Peer Review	22%	
MS Edit	10%	
Typeset	10% - 18%	
Derive e-materials		New Cost
Print and Mail	24%	
Other	34%	
Manage e-services		New Cost

Table Five

Process Analysis for Electronic Publishing Model Two

- Editorial
  - Peer Review
  - Data conversion to SGML
  - MS Edit in SGML
  - Derive e-materials from SGML
- Typeset from SGML
- Print and Mail
- Other
- Manage e-services

Table Six

Cost Analysis for Electronic Publishing Model Two

Editorial	
Peer Review	
Data conversion to SGML	*Additional Cost*
MS Edit in SGML	*Additional Cost*
Derive e-materials from SGML	*New Cost, less than Model One*
Typeset from SGML	*Reduced Cost*
Print and Mail	
Other	
Manage e-services	*New Cost*

……

For additional information about the conference, or The Andrew W. Mellon Foundation's scholarly communication initiatives, please contact Richard Ekman. For additional information about ARL or this web site contact Patricia Brennan, ARL Program Officer at (202) 296-2296.

Return to Office of Scholarly Communication Home Page

ARL Home ARL Scholarly Communication and Technology Home Page

© Association of Research Libraries, Washington, DC
Web Design by Angelo F. Cruz
Maintained by ARL Web Administrator
Last Modified: August 29, 1997

Scholarly Communication and Technology

Conference Organized by The Andrew W. Mellon Foundation

at Emory University
April 24-25, 1997

Copyright © of the papers on this site are held by the individual authors or The Andrew W. Mellon Foundation. Permission is granted to reproduce and distribute copies of these works for nonprofit educational or library purposes, provided that the author, source, and copyright notice are included on each copy. For commercial use, please contact Richard Ekman at the The Andrew W. Mellon Foundation.

Session #2 The Evolution of Journals

The Future of Electronic Journals

Hal R. Varian, Dean
School of Information, Management and Systems
University of California at Berkeley

It is widely expected that a great deal of scholarly communication will move to an electronic format. The Internet offers much lower cost of reproduction and distribution than print, the scholarly community has excellent connectivity, and the current system of journal pricing seems to be too expensive. Each of these factors are helping push journals from paper to electronic media.

In this paper I want to speculate about the impact this movement will have on the *form* scholarly communication. How will electronic journals evolve?

Each new medium has started by emulating the medium it replaced. Eventually the capabilities added by the new medium allow it to evolve in innovative, and often surprising, ways. Alexander Graham Bell thought that the telephone would be used to broadcast music into homes. Thomas Edison thought that recordings would be mostly of speech rather than music.

Marconi thought that radio's most common use would be two-way communication rather than broadcast.

The first use of the Internet for academic communication has been as as a replacement for the printed page. But there are obviously many more possibilities.

1. Demand and supply

In order to understand how journals might evolve, it is helpful to start with an understanding of the demand and supply for scholarly communication today.

1.1 Supply of scholarly communication

The academic reward system is structured to encourage the production of ideas. It does this by rewarding the production and dissemination of ``good'' ideas---ideas that are widely read and acknowledged.

Scholarly publications are produced by researchers as part of their jobs. At most universities and research organizations, publication counts significantly towards salary and job security (e.g., tenure.) All publications are not created equally: competition for space in top-ranked journals is intense.

The demand for space in those journals is intense because they are highly visible and widely read. Publication in a top flight journal is an important measure of visibility. In some fields, citation data has become an important observable proxy for ``impact.'' Citations are a way of proving that the articles that one publishes are, in fact, read.

1.2 Demand for scholarly communication

Scholarly communication also serves as an input to academic research. It is important to know what other researchers in your area are doing so as to improve your own work and to avoid duplicating their work. Hence, scholars generally want access to a broad range of academic journals.

The ability of universities to attract top-flight researchers depends on the size of the collection of the library. Threats to cancel journal subscriptions are met with cries of outrage by faculty.

1.3 The production of academic journals

[Tenopir and King(1996)] have provided a comprehensive overview of the economics of journal production. According to their estimates, the ``first-copy'' costs of an academic article are between $2,000 and $4,000. The bulk of these costs are labor costs, mostly clerical costs for managing the submission, review, editing, typesetting and setup costs.

The marginal cost of printing and mailing an *issue* of a journal is on the order of $3. A special-purpose, nontechnical academic journal that publishes 4 issues per year with 10 articles each issue would have fixed costs of about $120,000. The variable costs of printing and mailing would be about $12 per year. Such a journal might have a subscriber list of about 600, which leads to a break-even price of $212. Of course, many journals of this size are sold by for-profit firms and the actual prices may be much higher: prices of $600 or more are not uncommon for journals of this nature.

If the variable costs of printing and shipping were eliminated, the breakeven price would fall to $200. This illustrates the following point: fixed costs dominate the production of academic journals; reduction in printing and distribution costs due to electronic distribution will have negligible effect on breakeven prices.

Of course, if many new journals are produced and distributed electronically the resulting competition may chip away at the $600 monopoly prices. But if these new journals use the same manuscript-handling processes the $200 cost-per-subscription will remain the effective floor to journal prices.

2. Other costs

There are two other costs that should be mentioned. First is the cost of archiving. [Cooper(1989)] estimates that the present value of the storage cost of a single issue of a journal to a typical library is between $25 and $40.

Another interesting figure is yearly cost-per-article read. This varies widely by field, but we can offer a few order-of-magnitude guesses. According to a chart in [Lesk(1997)], p 218, 22% of scientific papers published in 1984 were not cited in the ensuing 10-year period. The figure rises to 48% for social science papers, and a remarkable 93% for humanities papers!

[Odlyzko(1997)] estimates that the cost per reader of a mathematical article may be on the order of $200. By comparison, the director of a major medical library has told me that his policy is to cancel journals for which the cost per article read appears to be over $50.

It is not commonly appreciated that one of the major impacts of online publication is that use can be easily and precisely monitored. Will academic administrators really pay subscription rates implying costs per reading of several hundred dollars?

3. Re-engineering journal production

It seems clear that reduction in the costs of academic communication can only be achieved by re-engineering the manuscript handling process. Here I use re-engineering in both its original sense---rethinking the process---and its popular sense---reducing labor costs.

The current process of manuscript handling is not particularly mysterious. The *American Economic Review* works something like this. The author sends 3 paper copies of an article to the main office in Princeton. The editor assigns each manuscript to a co-editor, based on the topic of the manuscript and the expertise of the co-editor. (The editor also reviews manuscripts in his own area of expertise.) The editor is assisted in these tasks by a staff of 2-3 FTE clerical workers.

The manuscript arrive in the co-editor's office who assigns them to two or more reviewers. The co-editor is assisted in this task by a half-time clerical worker. After some nudging, the referees usually report back and the co-editor makes a decision about whether the article merits publication. At the *AER* about 12% of the submitted articles are accepted

Typically the author revises accepted articles for both content and form, and the article is again sent to the referees for further review. In most cases the article is then accepted and sent to the main office for further processing. At the main office, the article is copyedited and further prepared for publication. It is then sent to be typeset. The proof sheets are sent to the author for checking. After corrections are made, the article is sent to the production facilities where it is printed, bound, and mailed.

Much of the cost in this process is the cost of coordinating the communication: the author sends the paper to the editor, the editor sends it to the co-editor, the co-editor sends it to referees, etc. These costs require postage and time, but most importantly they require coordination. This is the role played by the clerical assistants.

Universal use of electronic mail could undoubtedly save significant costs in this component of the publication process. The major enabling technology are be standards for document representation (e.g., Microsoft Word, PostScript, SGML, etc.) and multi-media email.

[Revelt(1996)] sampled Internet working paper sites and prepared a summary table. According to his survey, PostScript and PDF are the most popular formats for eprints with TeX being common in technical areas and HTML for non-technical areas. It is likely that standardization on 2-3 formats would be adequate for most authors and readers. My personal recommendation would be to standardize on Adobe PDF since it is readily available, flexible and inexpensive.

With respect to email, the market seems to be rapidly converging to MIME as a standard for email inclusion; I expect this convergence to be complete within a year or two.

This means that the standards are essentially in place to move to electronic document management during the editorial and refereeing process. Obviously new practices would have to be developed to ensure security and document integrity. Systems for timestamping documents such as Electronic Postmarks are readily available; the main barrier to their adoption is training necessary for their use.

4. Impact of re-engineering

If all articles were submitted and distributed electronically, I would guess that the costs of the editorial process would drop by a factor of 50% due to the reduction in clerical labor costs, postage, photocopying, etc.

Once the manuscript was accepted for publication, it would still have to be copyedited and converted to a uniform style. In most academic publishing copyediting is rather light, but there are exceptions. Conversion to a uniform style is still rather expensive due to the idiosyncrasies of authors' wordprocessing systems and writing habits.

It is *possible* that journals could distribute electronic style sheets that would help authors achieve a uniform style, but experience thus far has not given great reason for optimism on this front. Journals that accept electronic submissions report significant costs in conversion to a uniform style.

One question that should be taken seriously is whether these conversion costs for uniform style are worth it. Typesetting costs are about $15-$25 per page for moderately technical material. Markup costs probably require 2-3 hours of a copyeditor's time. This means that preparation costs for a 20-page article are on the order of $500. If a hundred people read the article, is the uniform style worth $5 apiece to them? Or, more to the point, if 10 people read the article is the uniform style worth $50 apiece?

The advent of desktop publishing dramatically reduced the cost of small-scale publication. But it is not obvious that the average quality of published documents went up. The earlier movement from hard type to digital typography had the same impact. As [Knuth(1979)] observes, digitally typeset documents cost less but had lower quality than handset documents.

My own guess about this benefit-cost tradeoff is that the quality from professional formatted documents isn't worth the cost for material that is only read by small numbers of individuals. The larger the audience, the more beneficial and cost-effective formatting becomes. This may suggest a two-tiered approach: articles that are formatted by authors are published very inexpensively. Of these, the ``classics'' can be ``reprinted'' in professionally designed formats.

A further issue arises in some subjects. Author-formatted documents may be adequate for reading, but they are not adequate for archiving. It is very useful to be able to search and manipulate subcomponents of an article such as abstracts and references. This means that the article must be formatted in a way that these subcomponents can be identified. Standard Generalized Markup Language (SGML) allows for such formatting, but it is rather unlikely that it could be used by most authors, at least using tools available today.

The benefits from structured markup are significant, but it is also quite costly so the benefit-cost tradeoff is far from clear. We return to this point below.

In summary, re-engineering the manuscript handling process by moving to electronic submission and review may save close to half of the first-copy costs of journal production. If we take the $2,000 first-copy costs per article as representative, this moves the first-copy costs to about $1,000. Moving the formatting responsibility to authors would reduce quality, but would also save even more on first-copy costs. For journals with small readership this tradeoff may be worth it. Indeed, many humanities journals have moved to online publication for reasons of reduced cost.

[Odlyzko(1997)] estimates that the cost of [Ginsparg(1996)]'s electronic preprint server is about between $5 and $75 per paper These papers are formatted entirely by the authors (mostly using TeX) and are not refereed. Creation and electronic distribution of scholarly work can be very inexpensive; one has to wonder whether the value added by traditional publishing practices is really worth it.

5. Electronic distribution

Up until now we have only considered the costs of preparing the manuscript for publication. If the material were subsequently distributed electronically there would be further savings. We can classify these as follows:

- Shelf-space savings to libraries. As we've seen these could be on the order of $35 per volume in present value. However, electronic archiving is not free. Running a Web server or creating a CD is costly. Even more costly is updating the media. Books that are hundreds of years old can easily be read today. Floppy disks that are 10 years old may be unreadable due to obsolete storage media or formatting. Electronic archives will need to be backed up, transported to new media and translated. All of these activities are costly. (Of course, traditional libraries are also costly; the ARL estimates this cost to be on the order of $12,000 per faculty member per year. Electronic documents will undoubtedly reduce many of the traditional library costs once it is fully implemented.)
- Monitoring. As mentioned above, it is much easier to monitor the use of electronic media. Since the primary point of the editorial and refereeing process is to economize on readers' attention, it should be very useful to have some feedback on whether articles are actually read. This would help make more rational decisions about journal acquisition, faculty retention, and other critical resource allocation issues.
- Search. It is much easier to search electronic media. References can be immediate displayed using hyperlinks. Both forward and reverse bibliographic searches can be done using online materials, which should greatly aid literature.
- Supporting materials. There are very small incremental costs to storing longer documents so it is easy to include data sets, images, detailed analyzes, simulations, etc. that can improve scientific communication.

5.1 Chickens and eggs

The big issue facing those who want to publish an electronic journal is how to get the ball rolling. People will publish in electronic journals when there are lots of readers; people will read electronic journals when there is lots of high-quality material published there.

This kind of ``chicken and egg'' problem is known in economics as a ``network externalities'' problem. We say a good (such as an electronic journal) exhibits network externalities if an individual's value for the product depend on how many other people use it. Telephones, faxes, and email all exhibit network externalities. Electronic journals exhibit a kind of indirect form of network externalities since the readers' value depends on how many authors publish in the journal and the number of authors who publish depend on how many readers there are.

There are several ways around this problem, most of which involve discounts for initial purchasers. You can give the journal away for a while, and eventually charge for it, as the *Wall Street Journal* has done. You can pay authors to publish in it, as the *Bell Journal of Economics* did when it started. It is important to realize that the payment doesn't have to be a monetary one. A very attractive form of payment is to offer ``prizes'' for the best articles published each year in the journal. The prizes can offer a nominal amount of money, but the real value is being able to list such a prize on your vitae. In order to be credible, such prizes should be juried and promoted widely. This may be a very nice way to overcome young authors' reluctance to publish in electronic journals.

6. When everything is electronic

Let us now speculate a bit about what will happen when all academic publication is electronic. I suggest that 1) publications will have much more general forms; 2) new filtering and refereeing mechanisms will be used; 3) archiving and standardization will remain a problem.

6.1 Document structure

The fundamental problem with specialized academic communication is that it is specialized. The number of readers of many academic publications is less than 100. Despite these small numbers, the academic undertaking may still be worthwhile. Progress in academic research comes by dividing problems up into small pieces and investigating these pieces in depth. Painstaking examination of minute topics provides the building blocks for grand theories.

However, there is much to be said for the viewpoint that academic research may be excessively narrow. It is said that a ghost named ``Pedro'' haunts the bell tower at Berkeley. The undergrads make offerings to Pedro at the Campanile on the evening before the exam. Pedro, it is said, was a graduate student in linguistics who wanted to right his thesis on Sanskrit. In fact, it was a thesis about one word in Sanskrit. And, it was not just one word, but in fact was on one of this word's forms in one of the particularly obscure declensions of Sanskrit. Alas, his thesis committee rejected Pedro's topic as ``too broad.''

However, the narrowness of academic publication is not entirely due to the process of research, but is also due to the costs of publication. Editors encourage short articles, partly to save on publication costs, but mostly to save on the attention costs of the readers. *Physics Letters* is widely read because the articles are required to be short. But one way authors achieve the required brevity is remove all ``unnecessary'' words ... such as conjunctions, prepositions, and articles.

Electronic publication eliminates the *physical* costs of length, but not the attention costs. Brevity will still be a virtue for some readers; depth will be a virtue for others. Electronic publication allows for mass customization of articles, much like the famous ``inverted triangle'' in journalism: there can be a one-paragraph abstract, a one-page executive summary, a four-page overview, a 20-page article, and a 50-page appendix. User interfaces can be devised to read this ``stretchtext.''

Some of these textual components can be targeted towards generalists in a field, some towards specialists. It is even possible that some components could be directed towards readers who are *outside* the academic specialty represented.

This possibility for variable-depth documents that can have multiple representations is very exciting. Well-written articles could appeal to both specialists and to those outside the specialty. The curse of the small audience could be overcome if the full flexibility of electronic publication were exploited.

7. Filtering costs

As I noted earlier, one of the critical functions of the academic publishing system is to filter. Work cannot be cumulative unless authors have some faith that prior literature is accurate. Peer review helps ensure that work meets appropriate standards for publication.

There is a recognized pecking order among journals, with high-quality journals in each discipline having a reputation for being more selective than others. This pecking order helps researchers focus their attention on areas that are thought by their profession to be particularly important.

In the last 25 years many new journals have been introduced, with the majority coming from the private sector. Nowadays almost anything can be published *somewhere* ... the only issue is where. Publication itself conveys little information about quality.

Many new journals are published by for-profit publishers. They make money by selling journals subscriptions, which generally means publishing more articles. But the value of peer review comes in being selective, a value almost diametrically opposed to increasing the output of published articles.

I mentioned above that one of the significant implications of electronic publication was that monitoring costs are much lower. It will be possible to tell with some certainty what is being read. This will allow for more accurate benefit/cost comparisons with respect to purchase decisions. But perhaps even more importantly it will allow for better evaluation of the significance of academic research.

Citation counts are often used as a measure of the impact of articles and journals. Studies in economics ([Laband and Piette(1994)]) indicate that most of the citations are to articles published in a few journals. More and more articles are being published, a smaller and smaller fraction of which are read. ([de Sola Pool(1983)].) It is not clear that the filtering function of peer review is working appropriately in the current environment.

Academic hiring and promotion policies contribute an additional complication. Researchers choose narrower and narrower specialties, making it more and more difficult to judge achievement locally. Outside letters of evaluation have become essentially worthless due to lack of privacy guarantees. The only thing left is the publication record and quantity of publication is easier to convey to non-experts than quality of publication.

The result is that young academics are encouraged to publish as much as possible in their first 5-6 years. Accurate measures of the impact of young researcher's work, such as citation counts, cannot be accumulated in this short a time period. One reform that would probably help matters significantly would be to put an upper limit on the number of papers submitted as part of tenure review. Rather than submitting everything published in the last 6 years, assistant professors could only submit their 5 best articles. This would, I suggest, lead to higher quality work, *and* higher quality decisions on the part of review boards.

7.1 Dimensions of filtering

If we currently suffer from a glut of information, electronic publication will only make matters worse. Reduced cost of publication and dissemination is likely to make more and more material available. This isn't necessarily bad; it simply means that the filtering tools will have to be improved.

I would argue that there are two dimensions on which journals filter papers: interest and correctness. The first thing a referee should ask is ``is this interesting?" If the paper is interesting, the next question should be ``is this correct?" Interest is relatively easy to judge; correctness is substantially more difficult. But there isn't much value in determining correctness if interest is lacking.

When publication was a costly activity, it was appropriate to evaluate papers prior to publication. Ideally only interesting and correct work manuscripts would undergo the expensive transformation of publication. Furthermore publication is a binary signal: either a manuscript is published or not.

Electronic publication is cheap. Essentially everything should be published, in the sense of being made available for download. The filtering process will take place *ex post*, so as to help users determine which articles are worth downloading and reading. As indicated above, the existing peer review system could simply be translated to this new medium. But the electronic media offer possibilities not easily accomplished in print media. Other models of filtering may be more effective and efficient.

7.2 A model for electronic publication

Allow me to sketch one such model for electronic publishing that is based on some of the considerations above. Obviously it is only one model; many models should and will be tried. However, I think that the model I suggest has some interesting features.

First, the journal assembles a board of editors. The function of the board is not only to provide a list of luminaries to grace the front cover of the journal; they will actually have to do some work.

Authors submit (electronic) papers to the journal. These papers have 3 parts: a one-paragraph abstract, a 5-page summary, and a 20-30 page conventional paper. The abstract is standard part of academic papers and needs no further discussion. The summary is modeled after the *Papers and Proceedings* issue of the *American Economic Review:* it should describe what question the author addresses, what methods were used to answer the question, and what the author found. The summary should be aimed at as broad an audience as possible. This summary would then be linked to the supporting evidence: mathematical proofs, econometric analysis, data sets, simulations, etc. The supporting evidence could be quite technical, and would probably end up being similar to current published papers in structure.

Initially, I imagine that authors would write a traditional paper and pull out parts of the introduction and conclusion to construct the summary section. This would be fine to get started, though I hope that the structure would evolve beyond this.

The submitted materials will be read by 2-3 members of the editorial board who will rate them with respect to how interesting they are. The editors will only be required to evaluate the 5-page summary, and are not necessarily responsible for evaluating the correctness of the entire article. There will be a common ``curve'' used by the editors; e.g., at most 10% of the articles would get the highest score. The ``Editorial score'' will be attached to the paper and it will be made available on the server. Editors will be anonymous; only the score will be made public.

Note that all papers will be accepted; the current ratings system of ``publish or not'' is replaced by a scale of (say) 1-5. Authors will be notified of the rating they received from the editors and they can withdraw the paper at this point if they choose to do so. However, once they agree that their paper be posted it cannot withdrawn (unless it is published elsewhere) although new versions of it can be posted and linked to the old one.

Subscribers to the journal can search all parts of the on-line papers. They can also ask to be notified by email of all papers that receive scores higher than some threshold or that contain certain keywords. When subscribers read a paper, they also score it with respect to its interest, and summary statistics of these scores are also (anonymously) attached to the paper.

Since all evaluations are available online, it would be possible to use them in quite creative ways. For example, I might be interested in seeing the ratings of all readers with whom my own judgments are closely correlated. (See [Konstan et al.(1997)Konstan, Miller, Maltz, Herlocker, Gordon, and Riedl] for elaboration of this scheme.) Or I might be interested in seeing all papers that were highly rated by Fellows of the Econometric Society or the Economic History Society.

This sort of ``social recommender'' system will help people to focus their attention on research that their peers---whomever they may be---find interesting. Papers that are deemed ``interesting'' can then be evaluated with respect to their correctness.

Authors can submit papers that comment upon or extend previous work. When they do so, they submit a paper in the ordinary way with links to the paper in question, as well as to other papers in this general area. This discussion of a topic forms a thread that can be traversed using standard software tools. See [Harnad(1995)] for more on this topic.

Papers that are widely read and commented upon will certainly be evaluated carefully for their correctness. Papers that aren't read may not be correct, but that presumably has low social cost. The length of the thread attached to a paper indicates how many people have (carefully) read it. If many people have read the paper and found it correct, a researcher may have some faith that the results satisfy conventional standards for scientific accuracy.

This model is unlike the conventional publishing model, but it addresses many of the same design considerations. The primary components are:

- Articles have varying depths, which allows them to appeal to a broad audience as well as satisfy specialists.
- Articles are rated first with respect to interest by a board of editors. Articles that are deemed highly interesting are then evaluated with respect to correctness.

- Readers can contribute to the evaluation process.
- The unit of academic discourse becomes a thread of discussion. Interesting articles that are closely read and evaluated can be assumed to be correct and therefor serve as basis for future work.

References

Cooper(1989)
Michael Cooper. A cost comparison of alternative book storage strategies. *Library Quarterly*, 59 (3): xx--xx, 1989.

de Sola Pool(1983)
Ithiel de Sola Pool. Tracking the flow of information. *Science*, 221 (4611): 609--613, 1983.

Ginsparg(1996)
Paul Ginsparg. Winners and losers in the global research village. Technical report, Los Alamos, 1996. http://xxx.lanl.gov/blurb/pg96unesco.html.

Harnad(1997)
Stephen Harnad. The paper house of cards (and why it's taking so long to collapse). *Ariadne*, 8, 1997. http://www.ukoln.ac.uk/ariadne/.

Harnad(1995)
Stevan Harnad. The post-Gutenberg galaxy: How to get there from here. *Times Higher Education Supplement*, 1995. http://cogsci.ecs.soton.ac.uk:80/~harnad/THES/thes.html.

Knuth(1979)
Donald Knuth. *TeX and Metafont: New Directions in Typesetting*. American Mathematical Society, Providence, RI, 1979.

Konstan et al.(1997)Konstan, Miller, Maltz, Herlocker, Gordon, and Riedl
Joseph A. Konstan, Bradley N. Miller, David Maltz, Jonathan L. Herlocker, Lee R. Gordon, and John Riedl. Grouplens : applying collaborative filtering to usenet news. *Communications of the ACM*, 40 (3): 77--87, 1997.

Laband and Piette(1994)
David N. Laband and Michael J. Piette. The relative impact of economics journals: 1970-1990. *Journal of Economic Literature*, 32 (2): 640--66, 1994.

Lesk(1997)
Michael Lesk. *Books, Bytes and Bucks: Practical Digital Libraries*. TBA, San Francisco, 1997.

Odlyzko(1997)
Andrew Odlyzko. The economics of electronic journals. Technical report, AT&T Labs, 1997.

Revelt(1996)
David Revelt. Electronic working paper standards. Technical report, UC Berkeley, 1996. http://alfred.sims.berkeley.edu/working-paper-standards.html.

Tenopir and King(1996)
Carol Tenopir and Donald W. King. Trends in scientific scholarly journal publishing. Technical report, School of Information Sciences, University of Tennesee, Knoxville, 1996.

...Varian
Research support from NSF grant SBR-9320481 is gratefully acknowledged.

...$212.
This neglects fixed costs such as marketing, overhead, etc., which would typically be included in a publisher's costs calculation.

...cost.
([Odlyzko(1997)] and [Harnad(1997)] have similar cost estimates.

Hal Varian
Wed Jun 11 09:38:05 PDT 1997

……

For additional information about the conference, or The Andrew W. Mellon Foundation's scholarly communication initiatives, please contact Richard Ekman. For additional information about ARL or this web site contact Patricia Brennan, ARL Program Officer at (202) 296-2296.

Return to Office of Scholarly Communication Home Page

ARL Home ARL Scholarly Communication and Technology Home Page

© Association of Research Libraries, Washington, DC
Web Design by Angelo F. Cruz
Maintained by ARL Web Administrator

Last Modified: August 29, 1997

Scholarly Communication and Technology

Conference Organized by The Andrew W. Mellon Foundation

at Emory University
April 24-25, 1997

Copyright © of the papers on this site are held by the individual authors or The Andrew W. Mellon Foundation. Permission is granted to reproduce and distribute copies of these works for nonprofit educational or library purposes, provided that the author, source, and copyright notice are included on each copy. For commercial use, please contact Richard Ekman at the The Andrew W. Mellon Foundation.

JSTOR: The Development of a Cost-Driven, Value-Based Pricing Model

Kevin M. Guthrie
Executive Director
JSTOR

• •

In the spring of 1996, when I was first approached to participate in this conference and was informed that the topic I was to address was pricing and user acceptance, I remember thinking it was quite a leap of faith, since JSTOR had neither a business model with prices, nor users. And we surely did not have user acceptance. Much has happened in a relatively short period of time, most notably the fact that JSTOR signed up 199 charter participants during the first three months of 1997. Our original projections were to have 50 to 75 participating institutions, so we are very encouraged to be off to such a good start.

The purpose of this brief case report is to summarize how JSTOR's economic model was developed, what we have learned along the way, and what we think the future challenges are likely to be. JSTOR is a work-in-progress, so it is not possible, nor would it be wise, to try to assert that we have done things "right." The jury is out, and will be for quite some time. My goal is only to describe our approach to this point in the hope that doing so will provide useful "experience" for others working in the field of scholarly communication. In providing this summary I will try not to stray far from the organizing topic assigned to me -- pricing and user acceptance -- but I think it is impossible to separate these issues from more general aspects of a not-for-profit's organizational strategy, and particularly its mission.

History

JSTOR began as a project of The Andrew W. Mellon Foundation designed to help libraries address growing and persistent space problems. Couldn't advances in technology help reduce the system-wide costs associated with storing commonly held materials like core academic journals? A decision was made to test a prototype system that would make the backfiles of core journals available in electronic form. Mellon Foundation staff signed up journal publishers in history and economics and, working through a grant to the University of Michigan, began to create a database with associated controlling software that was made available to several test site libraries. It became evident very soon that the concept was both extremely complicated to implement and that it held great promise.

JSTOR was established as an independent not-for-profit organization with its own Board of Trustees in August 1995. From the outset, JSTOR was given the charge to develop a financial plan that would allow it to become self-sustaining -- the Mellon Foundation was not going to subsidize the concept indefinitely. At the same time, JSTOR is fortunate to have had Mellon's initial support because enormous resources have been invested in getting the entity launched that never have to be paid back. Apart from the direct investments of funds in the development of software, production capacity, and mirror sites through grants to Michigan and Princeton, there were large investments of time and effort by Mellon Foundation staff. JSTOR has received, in effect, venture capital for which it need not produce an economic return. We have tried to translate these initial grants into lower prices for the services that we provide to JSTOR participants.

Defining "the Product"

Although JSTOR does not have to repay initial investments, it must have a mechanism to recover its ongoing costs. In developing a plan for cost recovery, our first step was to define exactly what it is that our "customers" would pay for -- what is the "product"? On the face of it, this step sounds simple, but it is anything but that, especially given the rate of change of technology affecting the Internet and World Wide Web. For example, those publishers reading this paper who are working to put current issues in electronic form will know that even choosing the display format can be extremely difficult. Should the display files be images or text? If text, should they be SGML, PDF, HTML, SGML-to-HTML converted in advance, SGML-to-HTML converted on the fly, or some combination of these or other choices? The format that is chosen has far-reaching implications for present and future software capabilities, charging mechanisms and user acceptance. It is easy to imagine how this decision alone can be paralyzing.

For nonprofit institutions like JSTOR, a key guidepost for making decisions of this type is the organization's mission. Nonprofits do not set out to maximize profits or shareholder wealth. In fact, they have been created to provide products or services that would not typically be made available by firms focused on maximizing profit. Consequently, not-for-profits cannot rely solely on quantitative approaches for decision-making, even when such decisions are quantitative or financial in nature. Without such tools, having a clearly defined mission and using it to inform decisions is essential.

A good example of how JSTOR has relied on its mission for decision-making is the question mentioned briefly above -- choosing an appropriate display format. We have decided to use a combination of images and text for delivery of the journal pages. We provide the images for display - so a user reads and can print a perfect replication of the original published page - and in the background we allow users to search the full text. This decision has been criticized by some people, but it is an appropriate approach for us, given the fact that our goal is to be a trusted archive and because JSTOR is now chiefly concerned with replicating previously published pages. There would be benefits to tagging the full text with SGML and delivering 100% corrected text files to our users, but because we also are committed to covering our costs, that approach is not practical. We are building a database of millions of pages and the effort required to do so is enormous. Digitizing even a single JSTOR title is a substantial undertaking. I have heard some people wonder why JSTOR is including "only" 100 journals in its first phase when other electronic journal initiatives are projecting hundreds, even thousands of journals. Presently, the 20 JSTOR journals that are available online have an average run of over 50 years. So any calculation about the effort required for converting a single title needs to be multiplied thirty to fifty times to be comparable to the effort required to publish an electronic version of a single year of a journal. That imposes very real constraints.

Having a clear understanding of our fundamental mission has also allowed us to remain flexible as we confront a rapidly evolving environment. It is a never-ending task trying to keep up with the technology. We work hard to remain open to change, and at the same time we are committed to using the appropriate technology to fulfill our objective - no more, no less. Progress can grind to a halt quickly when so much is unknown, and so much is changing, but our simple goal is to keep making progress. We recognize that by pushing forward relentlessly we will make some mistakes, but we are convinced that we cannot afford to stop moving if we are to build something meaningful in this dynamic environment.

So we established goals consistent with our mission and have made adjustments as we have gained experience. As mentioned previously, one of our fundamental goals is to serve as a trusted archive of the printed record. That means that output produced by the database has to be at least as good as the printed journals. A key determining factor in the quality of JSTOR printouts is the initial resolution at which the journal pages are scanned. Our original inclination was to scan pages at a resolution of 300 dots-per-inch (dpi). Anne Kenney[1] was a key advocate for scanning at 600 dpi when most people advised that 300 dpi was adequate and 600 dpi too expensive. Kenney made a strong case that scanning at 600 dpi is not just better than scanning at 300 dpi, but that, for pages comprised mainly of black-and-white text, there are rapidly diminishing perceivable improvements in the appearance of images scanned at resolutions beyond 600 dpi. It made sense, given the predominance of text in our database, to make the additional investment to gain the assurance that the images we were creating would continue to be acceptable even as technologies continued to improve. We are pleased that we made this choice; the quality of output now available from the JSTOR database is generally superior to a copy made from the original.

Another illustration of how it has been important for us to remain flexible concerns delivery of current issues. In the early days of JSTOR, several scholarly associations approached us with the idea that perhaps we could publish their current issues. The notion of providing scholars with access to the complete run of the journal - from the current issue back to the first issue - had (and has) enormous appeal. On the face of it, it seemed to make sense for JSTOR also to mount current issues in the database and we began to encourage associations to think about working with us to provide both current issues and the back files. It was soon evident, however, that this direction was not going to work for multi-title publishers. These publishers, some of which publish journals owned by others such as scholarly associations, justifiably regarded a JSTOR initiative on current issues to be competition. They were not about to provide the backfile of a journal to us only to risk that journal's owners turning to JSTOR for electronic publication of current and future issues. Again, we had to make adjustments. We are now committed to working with publishers of current issues to create linkages that will allow seamless searches between their data and the JSTOR archive, but we will not ourselves publish current issues.[2] If we are to have maximum positive impact on the scholarly community, we must provide a service that benefits not only libraries and scholars but also publishers of all types, commercial and not-for-profit, multi-title and single-title. It is part of having a system-wide perspective, something, which has been a central component of our approach from JSTOR's first days.

Determining Viability

Once we had framed the basic parameters of what we were going to offer, the key question we had to ask ourselves was whether it could be economically viable. Unfortunately, definitive answers to this question are probably never known in advance. The fact of the matter is that during their earliest phase, projects like JSTOR, even though they are not-for-profit, are still entrepreneurial ventures. They face almost all of the same risks as for-profit start-ups and the same tough questions must be asked before moving forward. Is there a revenue generating "market" for the service to be provided?[3] Does the enterprise have sufficient capital to fund up-front costs that will be incurred before adequate revenue can be generated? Is the market large enough to support the growth required to keep the entity vibrant?

Pursuing this analysis requires a complicated assessment of interrelated factors. What are the costs for operating the entity? That depends on how much "product" is sold. How much product can be sold, and what are the potential revenues? That depends on how it is priced. What should be the product's price? That depends on the costs of providing it. Because these factors are so closely related, none of them can be analyzed in isolation from the others; however, it is natural for a not-for-profit project focused on cost recovery to begin its assessment with the expense side of the ledger.

Defining the Costs

When the product or service is one that has not previously been offered, projecting potential costs is more art than science. Even if one has some experience providing a version of the product, as JSTOR had because of the Mellon initiative, one finds that the costs that have been incurred during the initial start-up period are irregular and unstable, and thus not reliable for projecting beyond that phase. Even now, with nearly 200 paying participants, we still have much to learn about what stable running costs are likely to be.

What we have learned is that our costs fall into basically six categories. They are:

1) *Production*: identifying, finding and preparing the complete run, defining indexing guidelines to inform a scanning sub-contractor, and performing quality control on the work of the scanning sub-contractor;

2) *Conversion*: scanning, OCR and inputting of index information to serve as the electronic table of contents (performed by a scanning sub-contractor);

3) *Storage and access*: maintaining the database (at a number of mirror sites), which involves continuous updating of hardware and systems software;

4) *Software development*: migrating the data to new platforms and systems and providing new capabilities and features to maximize its usefulness to scholars as technological capabilities evolve;

5) *User support*: providing adequate user help desk services for a growing user base;

6) *Administration and oversight*: managing the overall operations of the enterprise.

Some of these costs are one-time (capital) expenditures and some of them are on-going (operating) costs. For the most part, production and conversion (#1 and #2 above) are one-time costs. We hope that we are digitizing from the paper to the digital equivalent only once.[4] The costs in the other categories will be incurred regardless of whether new journals are added to the database and are thus a reflection of the ongoing costs of the enterprise.[5]

Because the most visible element of what JSTOR provides is the database of page images, many people tend to think that the cost of scanning is the only cost factor that needs to be considered. Although the scanning cost is relevant, it does not reflect the total cost of conversion for a database like JSTOR. In fact, scanning is not even the most expensive factor in the work done by our scanning contractor. During the conversion process, JSTOR's scanning vendor creates an electronic table of contents, which is just as costly as the scanning. In addition, because creating a text file suitable for searching requires manual intervention after running OCR software, that step has proven to be even more expensive than scanning. All told, the direct incremental costs of creating the three-part representation of a journal page in the JSTOR database (page image, electronic table-of-contents entry and text file) is approximately $0.75 to $1.00 per page.

Payments to the scanning bureau do not represent the complete production cost picture. Converting 100,000 pages per month requires a full-time staff to prepare the journals and to give the scanning bureau instructions to insure that table of contents and indexing entries are made correctly. At present production levels, these costs are approximately equal to the outlays made to the scanning bureau. On average then, JSTOR production costs approach $2.00 per page.

Other costs of operating JSTOR are less easily segregated into their respective functional "departments." Our present estimates are that once all of the 100 Phase I journals are available in the database, operating costs (independent of the one-time costs associated with production) will be approximately $2.5 million annually.

Defining Pricing

On the one hand, the obvious goal is to develop a pricing plan that will cover the $2.5 million in projected annual expenses plus whatever one-time production related expenses are incurred in converting the journals. This of course depends upon the rate at which the content is being digitized. For projects designed to recover costs by collecting fees from users, it is also important to assess whether the value of the service to be provided justifies the level of expenditures being projected.

In JSTOR's case, we evaluated the benefits to participants of providing a new and more convenient level of access to important scholarly material, while also attempting to calculate costs that might be saved by participants if JSTOR allowed them to free expensive shelf space. A central part of the reason for our founding was to provide a service to the scholarly community that would be both better and cheaper. That goal is one that remains to be tested with real data, but it can and will be tested as JSTOR and its participating institutions gain more experience.

Our initial survey of the research indicated that the cost of library shelf space filled by long runs of core journals was substantial. Using a methodology devised by Malcolm Getz at Vanderbilt and cost data assembled by Michael Cooper at UC-Berkeley, we estimated that the capital cost for storing a single volume ranged between $24 and $41.[6] It follows that storing the complete run of a journal published for 100 years costs the holding institution between $2,400 and $4,100. In addition, operating costs associated with the circulation of volumes are also significant and resources could be saved by substituting centrally managed electronic access to the material. Estimates of these costs for some of our original test site libraries indicated that costs in staff time for reshelving and other maintenance functions ranged from $45 annually for a core journal at a small college, to $180 per title at a large research library with heavy use. These estimates of savings do not take into account the long-term costs of preservation, or the time saved by users in finding articles of interest to them.

Although these estimates were not used to set prices, they did give us confidence that a pricing strategy could be developed that would offer good value for participating institutions. We set out to define more specifically the key components of the service we would offer and attempted to evaluate them both in the context of our mission and our cost framework. We found that deciding how to price an electronic product was extraordinarily complex and it was clear that there was no correct answer. This is by no means an exhaustive list, but some of the key factors that we weighed in our development of a pricing approach included:

* Will access be offered on pay-per-use model, or by subscription, or both?

* If by subscription, will the resource be delivered to individuals directly or via a campus site license?

* If by site license, how is the authorized community of users defined?

* Will there be price differentiation or a single price?

* If the price varies in some way for different types of licensees, what classifying approach will be used to make the determinations?

In making decisions we weighed the merits of various options by evaluating which seemed most consistent with JSTOR's fundamental objectives. For example, we wanted to provide the broadest possible access to JSTOR for the academic community. Because pricing on a pay-per-use model usually yields prices higher than the marginal cost of providing the product, we determined that this was not consistent with our goal. We did not want to force students and scholars to have to decide whether it would really be "worth it" to download and print an article. We wanted to encourage liberal searching, displaying and printing of the resource. In a similar vein, we concluded that it would be better to begin by offering institutional site licenses to participating institutions. We defined the site license broadly by establishing that authorized users would consist of all faculty staff and students of the institution, plus any walk-up patrons using library facilities.[7]

Another decision made to encourage broad access was our determination that different types of users should pay different prices for access. This is an approach called price differentiation, which is very common in industries with high fixed costs and low marginal costs (like airlines, telecommunications, etc.). We decided to pursue a value-based pricing approach that seeks to match the amount institutions would contribute to the value they would receive from participation. By offering different prices to different classes of institutions, we hoped to distribute the costs of operating JSTOR over as many institutions as possible, and in a fair way.

Once we had decided to offer a range of price levels, we had to select an objective method to place institutions into different price categories. We chose the Carnegie Classification of Institutions of Higher Education for pricing purposes. Our reason for choosing the Carnegie Classes was the fact that these groupings reflect the degree to which academic institutions are committed to research. Because the JSTOR database includes journals primarily used for scholarly research and would therefore be most highly valued by research institutions, the Carnegie Classes offered a rubric consistent with our aims. In addition to the Carnegie Classes, JSTOR factors in the FTE enrollment of each institution, making adjustments that move institutions with smaller enrollments into classes with lower price levels. We decided to break higher education institutions into four JSTOR sizes: Large, Medium, Small and Very Small.

Having established four pricing classes and a means for determining what institutions would fill them, we still had to set the prices themselves. In doing so, we thought both about the nature of our cost structure and the potential for revenue generation from the likely community of participants. We noted immediately that the nature of JSTOR's cost structure for converting a journal -- a large one-time conversion cost followed by smaller annual maintenance costs -- was matched by the nature of the costs incurred by libraries to hold the paper volumes. In the case of libraries holding journals, one-time or capital costs are reflected in the cost of land, building and shelves, while annual outlays are made for such items as circulation/reshelving, heat, light and electricity. We decided, therefore, to establish a pricing approach with two components: a one-time fee (which we called the Database Development Fee, or DDF) and a recurring fee (which we called the Annual Access Fee, or AAF).

But what should those prices be? As mentioned previously, the long-term goal was to recover $2.5 million in annual fees while also paying the one-time costs of converting the journals to digital formats. Because it was

BEST COPY AVAILABLE

impossible to model potential international interest in JSTOR, we limited our plan to U.S. higher education institutions. We conducted an assessment of the potential number of participants in each of our four pricing classifications. The number of U.S. higher education institutions in each category is shown in Table 1.

Table 1. Number of U.S. Higher Education Institutions by JSTOR Class

JSTOR Class	Number of Institutions
Large	176
Medium	589
Small	166
Very Small	471
Total	1,402

After thorough analysis of various combinations of prices, participation levels and cost assumptions, we arrived at a pricing plan we felt offered a reasonable chance of success. One other complicating aspect that arose as we developed the plan was how to offer a one-time price for a resource that was constantly growing. To deal with that problem, we defined our initial product, JSTOR-Phase I, as a database with the complete runs of a minimum of 100 titles in 10-15 fields. We promised that this database would be complete within three years. Prices for participation in JSTOR-Phase I are shown in Table 2.

Table 2. JSTOR Prices - Phase I

JSTOR Class	One-time Database Development Fee (DDF)	Annual Access Fee (AAF)
Large	$40,000	$5,000
Medium	30,000	4,000
Small	20,000	3,000
Very Small	10,000	2,000

These prices reflect the availability of the complete runs of 100 titles. That would mean that, for a Large institution, perpetual access to 80 years of The American Economic Review (1911-1991) would cost just $400 one-time and $50 per year. For a Small institution, it would be only $200 one-time and $30 per year. For comparison, consider that purchasing microfilm costs an order of magnitude more but offers far less convenient access. Also, if it proves to be possible to move copies to less expensive warehouses, or even to remove duplicate copies from library shelves, institutions will capture savings of some or all of the shelving and circulation costs outlined earlier in this paper. (For 80 volumes, that analysis projected capital costs of between $24 and $41 per volume, or $1,920 to $3,280 for an 80 volume run. Also, annual circulation costs were estimated as $180 per year for a Large institution.)

We purposely set our prices low in an effort to involve a maximum number of institutions in the endeavor. We are often asked how many participating institutions are needed for JSTOR to reach "breakeven." Because the total revenue generated will depend upon the distribution of participants in the various class sizes, there is no single number of libraries that must participate for JSTOR to reach a self-sustaining level of operations. Further, since our pricing has both one-time and recurring components, breakeven could be defined in a number of ways. One estimate would be to say that breakeven will be reached when revenues from annual access fees match non-production related annual operating expenditures (since the production related costs are primarily one-time). Although this is a useful guide, it is not totally accurate because, as mentioned previously, there are costs related to production that are very difficult to segregate from other expenses. Another approach would be to try to build an archiving endowment, and to set a target endowment size that would support the continuing costs of maintaining and migrating the Phase I archive, even if no additional journals or participants were added after the Phase I period. Our plan combines these two approaches. We believe it is important to match the sources of annual revenues to the nature of the

BEST COPY AVAILABLE

purposes for which they will be used. We require sufficient levels of annual inflows to cover the costs of making JSTOR available to users (user help desk, training, instruction, etc.). These should be collected by way of annual access fees from participants. There is also, however, the archiving function that JSTOR provides which is not directly attributable to any particular user. Like the role that libraries fill by keeping books on the shelves just in case they are needed, this is a public good. We must build a capital base to support the technological migration and other costs associated with this archiving function.

As with other aspects of our organizational plan, we remain open to making adjustments in pricing when it is fair, appropriate and does not put our viability at risk. One step we took was to offer a special charter discount for institutions that chose to participate in JSTOR prior to April 1, 1997. We felt it was appropriate to offer this discount in recognition of participants' willingness to support JSTOR in its earliest days. We also have made minor adjustments in the definitions of how Carnegie Classes are slotted into the JSTOR pricing categories. In our initial plan, we included all Carnegie Research (I and II) and Doctoral institutions (I and II) in the Large JSTOR category. As we spoke with librarians and administrators, it was clear that including Doctoral II institutions in this category was not appropriate. There proved to be a significant difference in the nature of these institutions and the resources they invest in research and so an adjustment was made to place them in the Medium class. In cases where we make adjustments of this nature, it is not for a single institution, but for all institutions that share a definable characteristic. In order to be fair, we do not believe in negotiating special deals.

There is a component of our pricing strategy that needs some explanation because it has been a disappointment to some people; that is, JSTOR's policy toward consortia. JSTOR's pricing plan was developed to distribute the costs of providing a shared resource among as many institutions as possible. The same forces that have encouraged the growth of consortia -- namely, the development of technologies to distribute information over networks -- are also what make JSTOR possible. It is not necessary to have materials shelved nearby in order to read them. A consequence of this fact is that marginal costs of distribution are low and economies of scale substantial. Those benefits have already been taken into account in JSTOR's economic model. In effect, JSTOR is itself a consortial enterprise that has attempted to spread its costs over as much of the community as possible. Offering further discounts to large groups of institutions would put JSTOR's viability, and with it the potential benefits to the scholarly community, at risk.

A second significant factor which prevents JSTOR from offering access through consortia at deep discounts is that the distribution of organizations in consortia is uneven and unstable. Many institutions are members of several consortia, while some are in none at all (although there are increasingly few of those remaining). If the consortial arrangements were more mature and there was a one-to-one relationship between the institutions in JSTOR's community and consortial groups, it might have been possible for JSTOR to build a plan that would distribute costs fairly across those groups. If, for example, every institution in the United States was a member of one of five separate consortia, a project like JSTOR could divide its costs by five and a fair contribution could be made by all. But there are not five consortia; there are hundreds. The patchwork of consortial affiliations is so complex that it is extremely difficult, if not impossible, to establish prices that will be regarded as fair by participants. JSTOR's commitment to share as much of what it learns with the scholarly community as possible requires that there be no special deals, that we be open about the contributions that institutions make and their reasons for making them. Our economic model would not be sustainable if two very similar institutions contributed different amounts simply because one was a member of a consortium that drove a harder bargain. Instead, we rely on a pricing unit which is easily defined and understood -- the individual institution. And we rely on a pricing gradient, the Carnegie Classification, which distributes those institutions objectively into groupings that are consistent with the nature and value of our resource.

Conclusion

BEST COPY AVAILABLE

The initial response to JSTOR's charter offer in the first three months of this year is a strong signal that JSTOR will be a valued resource for the research community; however, it is still far too early to comment further on "user acceptance." Tom Finholt's research (also presented at this conference) into usage at the test site libraries provides a first snapshot, but this picture was taken prior to there being any effort to increase awareness of JSTOR in the community and on the specific campuses. There is much to learn. By the conclusion of the 1997-1998 academic year, there will be more to say about whether the availability of JSTOR has any impact on the use of older journals. JSTOR is committed to tracking usage data both for libraries and publishers and to providing special software tools to enable users to create usage reports tailored to their own needs and interests. We will continue to keep the academic community informed as we learn more.

While we are encouraged by the positive reaction of the library community to JSTOR, we recognize that this good start has raised expectations and has created new challenges. In addition to reaching our 100-title goal before the end of 1999, trying to encourage the next 200 libraries to participate, and keeping up with changing technologies, we face other complex challenges, including how to make JSTOR available outside of the United States, and how to define future phases of JSTOR. Addressing these issues will require the development of new strategic plans and new economic and pricing models. In creating those plans, we know that we will continue to confront complicated choices. As we make decisions, we will remain focused on our mission, making adjustments to our plans as required to keep making progress in appropriate ways.

REFERENCES:

[1] Anne Kenney is the Associate Director of Preservation at the Cornell University Library. She has also authored a paper for this conference. It is entitled "Digital Image Quality: From Conversion to Presentation and Beyond."

[2] We did agree to work with three scholarly associations, the Ecological Society of America, the American Economic Association and the American Political Science Association to provide access to current issues through JSTOR. We stand by our commitments to these organizations, but our goal is to learn more about the technology required to make linkages between current issues and the archive, not to build the capability for JSTOR to become a publisher of current issues.

[3] In the not-for-profit context, that market need not consist solely of paying customers, it could include other types of indirect funders like government agencies or foundations.

[4] This is not precisely accurate. Not all of the production costs are one-time. We add another volume of each journal title to the database as each year passes, so there is an on-going element of the production costs, but they represent a small fraction of total production expenditures.

[5] There is a caveat here as well. Some of the administrative and overhead costs are higher because JSTOR is adding titles. Negotiating agreements with publishers is a time-consuming task, as is overseeing the

production operation converting 100,000 pages per month. It is not practical, however, to allocate exactly the portion of general administrative and other costs that pertain directly to production.

[6] For a more complete description of these estimates, see "JSTOR and the Economics of Scholarly Communication," a paper by William G. Bowen, which is available at http://www.mellon.org/jsesc.html.

[7] For a more complete description of the evolution in the development of JSTOR's library license terms, see "JSTOR: An IP Practitioner's Perspective," by Sarah E. Sully, D-Lib, January 1997.

• •

For additional information about the conference, or The Andrew W. Mellon Foundation's scholarly communication initiatives, please contact Richard Ekman. For additional information about ARL or this web site contact Patricia Brennan, ARL Program Officer at (202) 296-2296.

Return to Office of Scholarly Communication Home Page

ARL Home ARL Scholarly Communication and Technology Home Page

© Association of Research Libraries, Washington, DC

Web Design by Angelo F. Cruz
Maintained by ARL Web Administrator
Last Modified: August 14, 1997

BEST COPY AVAILABLE

Scholarly Communication and Technology

Conference Organized by The Andrew W. Mellon Foundation

at Emory University
April 24-25, 1997

Copyright © of the papers on this site are held by the individual authors or The Andrew W. Mellon Foundation. Permission is granted to reproduce and distribute copies of these works for nonprofit educational or library purposes, provided that the author, source, and copyright notice are included on each copy. For commercial use, please contact **Richard Ekman** at the The Andrew W. Mellon Foundation.

Session #3 Economics of Electronic Publishing: Journals Pricing and User Acceptance

The Effect of Price: Early Observations

Karen Hunter
Senior Vice President
Elsevier Science

. . . .

FINAL VERSION

THE EFFECT OF PRICE: EARLY OBSERVATIONS

Karen Hunter
k.hunter@elsevier.com

INTRODUCTION

Scientific journal publishers have very little commercial experience with electronic full text distribution and it is hard, if not impossible, to segregate the effect of pricing on user acceptance and behavior. Most experiments or trial offers have been without charge to the user. Most paid services have targeted institutional rather than individual buyers. Nevertheless, we can look at some of the known experiences and at ongoing and proposed experiments to get some sense of the interaction of pricing and acceptance and of the other factors, which seem to affect user behavior. We can also look at institutional buying concerns and pricing considerations.

IN THE BASIC PAPER WORLD...

Many journals have offered reduced prices to individuals. In the case of journals owned by societies or other organizations, there are generally further reductions in the prices for members. It is important to the society that members not only receive the lowest price but can clearly see that price as a benefit of membership. The price for members may be at marginal cost, particularly if (1) the size of the membership is large, (2) subscriptions are included as a part of the membership dues, and (3) there is advertising income to be gained from the presence of a large individual subscription base. One sees this commonly in clinical medical journals, where the presence of 15,000 or 30,000 or more individual subscribers leads to >$1 million in advertising income -- income which would be near zero without the individual subscription base. One can "afford" to sell the subscriptions at cost because of the advertising.

For many other journals, including most published by my company, there either are no individual rates or the number of individual subscribers is trivial. This is largely because the size of the journals, and therefore their prices, are sufficiently high (average $1,600) that it is difficult to set a price for individuals which would be attractive. Giving even a 50% reduction in price does not bring the journal into the price range that attracts individual purchasers.

One alternative is to offer a reduced rate for personal subscriptions to individuals affiliated with an institution which has a library subscription. This permits the individual rate to be lower, but it is still not a large source of subscriptions in paper. The price is still seen as high (*e.g.*, the journal *Gene* has an institutional price of $6,144 in 1997 and an associated personal rate of $533; the ratio is similar for *Earth and Planetary Sciences*, $2,333 for an institutional subscription, $150 for individuals affiliated with that institution.) This still draws only a very limited number of subscribers.

We have not recently (this decade) rigorously tested alternative pricing strategies for this type of paper arrangement nor talked with scientists to learn specifically why they have or have not responded to an offer. This reflects a view that there is only limited growth potential in paper distribution and that the take-up by individuals (if it is to happen) will be in an electronic world.

ALTERING SERVICES

There is some experience with free distribution, which may be relevant. Over the last decade we have developed a fairly large number of electronic and paper services designed to "alert" our readers to newly-published or soon-to-be-published information. These services take many forms, including lists of papers accepted for publication; current tables of contents; groupings of several journals in a discipline; single journal-specific alerts; inclusion of additional discipline-specific news items, etc. Some are mailed. Some are electronically broadcast. Others are electronically profiled and targeted to a specific individual's expressed interest. Finally, some are simply on our server and "pulled" on demand.

All are popular and all are sent only to users who have specifically said they want to receive these services. The electronic services are growing rapidly, but the desire for those which are paper-based continues. One even sees "claims" for missing issues, should a copy fail to arrive in the mail. What we take from this is that there is a demand for information about our publications -- the earlier the better -- and that so long as it is free *and* perceived as valuable, it will be welcomed. Note, however, that in the one case where, together with another publisher, we tried to increase the perceived value of an alerting service by adding more titles to the discipline cluster and adding some other services, there was noticeable resistance to paying a subscription for the service.

ELECTRONIC PRICING

In developing and pricing new electronic products and services, journal publishers may consider many factors, including (in random order):

- the cost of creating and maintaining the service;
- the possible effect of this product or service on other things you sell ("cannibalization" or substitution);
- the ability to actually implement the pricing (site or user community definitions, estimates of the anticipated usage or number of users, security systems)
- provision for price changes in future years
- what competitors are doing;
- the functionality actually being offered;
- the perceived value of the content and of the functionality;
- the planned product development path (in markets, functionality, content);
- the ability of the market to pay for the product or service;
- the values that the market will find attractive (e.g., price predictability or stability);
- the anticipated market penetration and growth in sales over time;
- the market behavior that you want to encourage;
- and, not inconsequentially, the effect on your total business if you fail with this product or service.

To make informed judgments, one has to build up experience and expertise. Pricing has long been an important strategic variable in the marketing mix for more mature electronic information players. They have more knowledge of how a market will react to new pricing models. For example, more than five years ago, one would see at an Information Industry Association meeting staff from business, financial and legal online services with titles such as Vice President, Pricing. Nothing comparable existed within the journal publishing industry. A price was set, take it or leave it, and there was little room for nuance or negotiation.

This is now changing. Many large journal publishers are actively involved in either negotiating pricing agreements or, under fixed terms, negotiating other aspects of the licensed arrangement which relate to the effective price being paid (such as number of users, number of simultaneous accesses, etc.) At Elsevier in 1996, we engaged consultants to make a rigorous study to assist us in developing pricing models for electronic subscriptions and other electronic services. What we found was that we could not construct algorithms to predict buying behavior in relation to price. That has not stopped us from trying to pursue more sophistication in pricing -- and indeed, we have now hired our own first full-time Director of Pricing -- but until we build up more experience, it still makes pricing decisions often a combination of tradition, strategic principle, gut-feeling and trial and error. There is, however, a view on the desired long-term position and how we want to get there.

Too often, some buyers argue that pricing should be based solely on cost (and often without understanding what goes into the cost). Therefore, there is a sometimes expressed a simplistic view that electronic journals are paper journals without the paper and postage and should be priced at a discount. That clearly is naive, overlooking all of the new, additional costs which go into creating innovative electronic products (as well as maintaining two product lines simultaneously). Indeed, if one were to price right now on simply the basis of cost, the price for electronic products would likely be prohibitively high.

It is equally doubtful if one can accurately determine the value added from electronic functionality and set prices based *exclusively* on the value, with the notion that as more functionality is added, the value -- therefore, the price -- can be automatically increased. Some value-based pricing is to be expected and is justified, but in this new electronic market there are also limited budgets and highly competitive forces, which keep prices in check. At the same time, it is not likely that the "content" side of the information industry will totally follow the PC hardware side -- i.e., that the prices will stay essentially flat, with more and more new goodies bundled in the product. Hardware is much more of a competitive commodity business.

Pricing components are now much more visible and subject to negotiation. In discussions with large accounts, it is assumed that there will be such negotiation. This is not necessarily a positive development for either publishers or libraries. One hopes that collectively we won't wind up making the purchase of electronic journals the painful equivalent of buying a car. ("How about some rust proofing and an extended warranty?")

There is and will continue to be active market feedback and participation on pricing. The most obvious is a refusal to buy, either because the price is too high, the price-value trade-off is not there, or because of other terms and conditions associated with the deal. Other feedback will come via negotiation and public market debates. Over time, electronic journal pricing will begin to settle into well-understood patterns and principles. At the moment, however, there are almost as many definitions and models as there are publishers and intermediaries. One need only note the recent discussions on the e-list on library licensing moderated by Ann Okerson of Yale University to understand that we are all in the early stages of these processes. An early 1997 posting gave a rather lengthy list of pricing permutations.

END USER PURCHASING

If one talks of pricing and "user acceptance", an immediate question is: who is the user? Is it the end user or is it the person paying the bill, if they are not one and the same? One presumes the intention was to reflect the judgments made by end users when those end users are also the ones bearing the economic consequences of their decisions. In academic information purchasing (as with consumer purchasing), the end user has traditionally been shielded from the full cost (often any cost) of information. Just as newspapers and magazine costs are heavily subsidized by advertising, and radio and television revenues (excluding cable) are totally paid by advertisers, so do academic journal users benefit from the library as the purchasing agent.

In connection with the design of its new Web journal database and host service, *ScienceDirect*^{TM}, Elsevier Science in 1996 held a number of focus groups with scientists in the U.S. and the UK. Among the questions asked was the amount of money currently spent personally (including from grant funds) annually on the acquisition of information resources. The number was consistently below \$500 and was generally between \$250 and \$400, often including society dues, which provided journal subscriptions as part of the dues. There was almost no willingness to spend more money, and there was a consistent expectation that the library would continue to be the provider of services, including new electronic services.

This is consistent with the results of several years of direct sales of documents through the (now) Knight-Ridder CARL UnCover service. When it introduced its service a few years ago, UnCover had expected to have about 50% of the orders coming directly from individuals, billed to their credit cards. In fact, as reported by Martha Whitaker of CARL during the 1997 annual meeting of the Association of American Publishers, Professional/Scholarly Publishing Division in February, the number has stayed at about 20% (of a modestly growing total business).

From their side, libraries are concerned that the user has little or no appreciation of the cost to the library of fulfilling their users' requests. In two private discussions in February of 1997, academic librarians told me of their frustration when interlibrary loan requests are made, the articles procured and the requesters notified, but then the articles are not picked up. There is a sense that this service is "free", even though it is well-documented (via a Mellon study) that the cost is now more than \$30 per ILL transaction.

In this context, discussions with some academic librarians about the introduction of electronic journal services have not always brought the expected reactions. It had been our starting premise that electronic journals should mimic paper journals in certain ways, most notably that once you have paid the subscription, then you have unlimited use within the authorized user community. However, one large library consortium negotiator has taken the position that he is not so sure that is desirable, as it would be better to start educating users that information has a cost attached to it.

Similarly, other librarians have expressed concern about online facilities which permit users to acquire individual articles on a transactional basis from non-subscribed titles (*e.g.*, in a service such as *ScienceDirect(TM)*). While the facilities may be in place to bill the end user directly, the librarians believe the users will not be willing to pay the likely prices (\$15-25). Yet, if the library is billed for everything, either the cost will run up quickly or any prepaid quota of articles will be used equally rapidly. The notion that was suggested was to find some way to make a nominal personal charge of perhaps $1 or $2 or $3 per transaction. It was the librarians' belief that such a charge would be enough to make the user stop and think before ordering something that would result in a much larger ultimate charge to the library.

The concern that demand could swamp the system if unregulated is one that would be interesting to test on a large scale. While there have been some experiments which I will describe further below, we have not yet had sufficient experience to generalize. Journal users are, presumably, different from America Online customers, who so infamously swamped the network in December 1996 when pricing was changed from time-based to unlimited use for $19.95 per month. Students, faculty and other researchers read journals for professional business purposes and generally try to read as little as possible -- *i.e.*, to be efficient in combing and reviewing the literature and not to read more and more without restraint. The job of a good electronic system is to increase that efficiency by providing tools to sift the relevant from the rest.

It is interesting to note that in a paper environment, the self-described "king of cancellations," Chuck Hamaker of Louisiana State University, reported during the 1997 mid-winter ALA meeting that he had canceled $738,885 worth of subscriptions between 1986 and 1996 and substituted free, library-sanctioned, commercial document delivery services. The cost to the library has been a fraction of what the subscription cost would have been. He now has about 900 faculty and students who have profiles with the document deliverer (UnCover) and who order directly, on an unmediated basis, with the library getting the bill. He would like to see that number increase (as there are 5000 faculty and students who would qualify). It will be interesting to see if the same pattern will occur if the article is physically available on the screen and the charge is incurred as a result of downloading. Will the decision to print be greater (because it is immediate and easy) than to order from a document delivery service?

This highlights one of the issues surrounding transactional selling: how much information is enough to have before ordering in order to insure that the article being ordered will be useful? Within the *ScienceDirect(TM)* environment we hope to answer this by creating services specifically for individual purchase which offer the user an article snapshot or summary (SummaryPlus^{SM}), which includes much more than the usual information about the article (*e.g.*, it includes all tables and graphs and all references). From the summary the user can make a much more informed decision about whether to purchase the full article.

TULIP (THE UNIVERSITY LICENSING PROGRAM)

Elsevier Science has been working toward the electronic delivery of its journals for nearly two decades. Its early discussions with other publishers about what became ADONIS started in 1979. Throughout the 1990s there have been a number of large and small programs, some experimental, some commercial. Each has given us some knowledge of user behavior in response to price, although in some cases the "user" is the institution rather than the end user. The largest experimental program was TULIP.

TULIP was a five year experimental program (1991-1995) in which Elsevier partnered with nine leading U.S. universities (including all of the universities within the University of California system) to test desktop delivery of electronic journals. The core of the experiment was the delivery of initially 43, later an additional optional 40, journals in materials science. The files were bitmapped (TIFF) format, with searchable ASCII headers and unedited, OCR-generated

ASCII full text. The universities received the files and mounted them locally, using a variety of hardware and software configurations. The notion was to integrate or otherwise present the journals consistently with the way other information was offered on campus networks. No two institutions used the same approach and the extensive learning gained has been summarized in a final report (available on request).

For the purposes of this paper, there are only a few relevant observations. First, the libraries (through whom the experiment was generally managed) generally chose a conservative approach in a number of discretionary areas. For example, while there was a document delivery option for titles not subscribed to (for each library received the electronic counterparts of their paper subscriptions), no one opted to do this. Similarly, the full electronic versions of non-subscribed titles were offered at a highly discounted rate (30% of list) but essentially found no takers. The most frequently expressed view was that a decision had been made at some time not to subscribe to the title, so its availability even at a reduced rate was not a good purchasing decision.

Second, one of the initial goals of this experiment was to explore economic issues. While the other goals (technology testing and evaluating user behavior) were well-explored, the economic side was less developed. That was perhaps a failure in the initial expectations and in the experimental design. From our side as publisher, we were anxious to try out different distribution models on campus, including models where there would be at least some charge for access. However, this was never set as a requirement, nor were individual institutions assigned to different economic tests. And, in the end, all opted to make no charges for access. This was entirely understandable, both because of the local campus cultures and the other issues to be dealt with in simply getting the service up and running, promoting it to users, etc. However, it did mean that we never gathered any data in this area.

From the universities' side, there was a hope that there would be more progress toward developing new subscription models. We did have a number of serious discussions, but again not as much was achieved as might have been hoped for if the notion was a radical change in the paradigm. I think everyone is now more experienced and realizes that these things are complex and take time.

Finally, the other relevant finding from the TULIP experiment is that use was very heavily related to the (lack of) perceived critical mass. Offering journals to the desktop is only valuable if it is the right journals and they are supplied on a timely basis. Timeliness was compromised because the electronic files were produced after the paper -- a necessity at the time but not how we (or other publishers) are currently proceeding. Critical mass was also compromised because, although there was a great deal of material delivered (11 GB per year), materials science is a very broad discipline and the number of journals relevant for any one researcher was still limited. If the set included "the" journal or one of the key journals a researcher (or more likely, graduate student) needed, use was high. Otherwise, there was not enough to remind users to return regularly to the system. And this is when there was no charge for use.

ELSEVIER SCIENCE EXPERIENCES WITH COMMERCIAL ELECTRONIC JOURNALS

- *Elsevier Electronic Subscriptions*

The single largest Elsevier program of commercial electronic delivery is the Elsevier Electronic Subscriptions (EES) program. This is the commercial extension of the TULIP program to all 1,100 Elsevier primary and review journals. The licensing negotiations are exclusively with institutions, which receive the journal files and mount them on their local network. The license gives the library unlimited use of the files within their authorized user community. As far as we are aware, academic libraries are not charging their patrons for their use of the files, so there is no data relating user acceptance to price. At least one corporate library charges use back to departments, but this is consistent with its practice for all of its services and has not affected use as far as is known.

If you broaden "user" to include the paying institution, as discussed above, then there is clearly a relation between pricing and "user" acceptance. If we can't reach an agreement on price in license negotiations, there is no deal. And it is a negotiation. The desire from the libraries is often for price predictability over a multi-year period. Because prices are subject to both annual price increases and the fluctuation of the dollar, there can be dramatic changes from year to year. For many institutions, the deal is much more "acceptable" if these increases are fixed in advance.

The absolute price is also, of course, an issue. There is little money available and pricing of electronic products at a high price will result in a reluctant end to discussions. Discussions are both easier and more complicated with consortia. It is easier to make the deal a winning situation for the members of the consortium (with virtually all members getting access to some titles which they had previously not had), but it is more complicated because of the number of parties who have to sign off on the transaction.

Finally, for a product such as EES, the total cost to the subscribing institution goes beyond what is paid to Elsevier as publisher. There is the cost of the hardware and software to store and run the system locally, the staff needed to update and maintain the system, local marketing and training time, etc. It is part of the sales process on the Elsevier side to explain these costs to the subscribing institution, as it is not in our interest or theirs underestimate the necessary effort, only to have it become clear during implementation. To date, our library customers have appreciated that approach.

- *Immunology Today Online (ITO)*

Immunology Today is one of the world's leading review journals, with an ISI impact factor of over 24. It is a monthly magazine-like title, with a wide individual and institutional subscription base. (The Elsevier review magazines are the exception to the rule in having significant individual subscriptions.) In 1994 its publishing staff decided it was a good title to launch also in an electronic version. They worked with OCLC to make it a part of the OCLC Electronic Journals Online collection, initially offered via proprietary Guidon software and launched in January, 1995.

As with other journals then and now making their initial online appearance, the first period of use was without charge. A testbed developed of about 5.0% of the individual subscribers to the paper version and 3.0% of the library subscribers. In time, there was a conversion to paid subscriptions, with the price for the combined paper and electronic personal subscriptions being 125% of the paper price. (You did not have to have both paper and electronic -- but only 3 people chose to take electronic only.) At the time OCLC ended the service at the end of 1996 and we began the process of moving subscribers to a similar Web version of our own, the paid subscription level for individuals was up to about 7.0% of the individual subscribers and 0.3% of the institutional subscribers.

The poor take-up by libraries was not really a surprise. At the beginning, libraries did not know how to evaluate or offer to patrons a single electronic journal subscription, as opposed to a database of journals. (There is a steady improvement in this area, provoked in part by the journals -- notably *The Journal of Biological Chemistry* or JBC -- offered via High Wire Press.) How do you let people know it is available? How and where is it available? And is a review journal -- even a very popular review journal -- the place to start? It apparently seemed like more trouble than it was worth to many librarians.

In talking with the individual subscribers -- and those who did not subscribe -- it was clear that price was not a significant factor in their decisions. The functionality of the electronic version was the selling point. It has features which are not in the paper and is, of course, fully searchable. That means the value was in part in efficiency -- the ease with which one found that article that you recalled reading six months ago but don't remember the author or precise month or search for information on a topic newly of interest. The electronic version is a complement to the paper, not a substitute. For those who chose not to subscribe, either they were deterred by the initial OCLC software (which had its problems) and may now be lured back via our Web version or they have not yet seen a value which will add to their satisfaction with paper. But it has not been a question of price.

- *Journal of the American College of Cardiology*

This project was somewhat different. This flagship journal is owned by a major society and has been published by Elsevier Science since its beginning in the early 1980s. In 1995, in consultation with the society Elsevier developed a CD-ROM version. The electronic design -- style, interface and access tools -- is quite good. The cost of the CD-ROM is relatively low ($295 for institutions, substantially less for members) and it includes not only the journal, but also five years of JACC abstracts, the abstracts from the annual meeting and one year (6 issues) of another publication entitled *ACC Current Reviews*.

But it has sold only modestly well. Libraries, again, resist CD-ROMs for individual journals (as opposed to journal collections). And the doctors have not found it a compelling purchase. Is it price per se? Or is it the notion of paying anything more, when the paper journal comes bundled as part of the membership dues? Or is there simply no set of well-defined benefits? Clearly, the perceived value to the user is not sufficient to cause many to reach for a credit card.

- *GeneCOMBIS, Earth and Planetary Sciences Online, etc.*

I mentioned above that for some paper journals we have personal rates for individuals at subscribing institutions. This model has been extended to Web products related to those paper journals. I mentioned above the journal *Gene*. In addition to the basic journal, *Gene*, we publish an electronic section called *GeneCOMBIS (for Computing for Molecular Biology Information Service)*, which is an electronic-first publication devoted to the computing problems that arise in molecular biology. It publishes its own new papers. The papers are also published in hard copy, but the electronic version includes hypertext links to programs, datasets, genetics databases and other software objects. *GeneCOMBIS* is sold to individuals for $75 per year, but only to those individuals whose institutions subscribe to *Gene*.

The same model is repeated with the electronic version of a leading earth sciences journal, *Earth and Planetary Science Letters*. The affiliated rate for the electronic version has been introduced in 1997, with a nominal list price of $90 and a 1/2 price offer for 1997 of $45. This provides online access to the journal and to extra material such as datasets for individuals affiliated with subscribing institutions.

It is too early to know whether this model will work. There certainly has been interest. In the case of *GeneCOMBIS,* ultimately its success will depend on the quality and volume of the papers it attracts. With *EPSL Online,* it will be the perceived value of the electronic version and its added information. In neither case is it expected that price will have a significant effect on subscriptions. What is more likely to happen is pressure to extend the subscriptions to individuals working outside institutions, which have the underlying paper subscriptions.

EXPERIENCES OF OTHERS

It is perhaps useful to note also some of the experiences of other publishers.

- *Red Sage experiment*

This experiment started in 1992 and ran through 1996. It was initially started by Springer-Verlag, the University of California at San Francisco and AT&T Bell Labs. Ultimately, several other publishers joined in and there were over 70 biomedical journals being delivered to the desktops of medical students and faculty at UCSF. As with TULIP, the experiment proved much harder to implement than had been originally hoped for. To the best of my knowledge, there were no user charges, so no data on the interplay of price and user acceptance. But what is notable is that there was greater critical mass of user-preferred titles among the Red Sage titles and, as a result, usage was very high. The horse will drink if brought to the right water.

- *Society CD-ROM options*

A second anecdote comes from discussions last year with a member of the staff of the American Institute of Physics. At least one of their affiliated member societies decided to offer members an option to receive their member subscriptions on CD-ROM rather than in paper, at the same price (*i.e.*, the amount allocated from their member dues). The numbers I recall are that over 1,500 members of the society took the option, finding that a more attractive alternative. One suspects that had they tried to sell the CD-ROM on top of the cost of the basic subscription, there would have been few takers. However, in this case if you ignored the initial investment to develop the CD, it saved the society money as well, as it *was* cheaper on the incremental cost basis to make and ship the CDs rather than print and mail the paper. In this case, the economics favored everyone.

- *BioMedNet*

The final observation relates to an electronic service that started last year called

BioMedNet. It is a "club" for life scientists, offering some full text journals, Medline, classified ads (the most frequently used service), marketplace features, news and other items. To date, membership is free. There are over 55,000 members and another 1000+ coming in each week. The site is totally underwritten at the moment by its investors, with an expectation of charging for membership at some later date but with the plan that principal revenues will come from advertising and a share of marketplace transactions. The observation here is that while the membership is growing steadily, usage is not yet high per registered member. There is a core of heavy users, but it is rather small (2-3%). So, again, behavior and acceptance is not a function of price but of perceived value. Is it worth my time to visit the site?

PEAK: THE NEXT EXPERIMENT

As was mentioned above, the aspect of the TULIP experiment that produced the least data was the economic evaluation. One of the TULIP partners was the University of Michigan, which is now also an Elsevier Electronic Subscription subscriber for all Elsevier journal titles. As part of our discussions with Michigan, we agreed to further controlled experimentation in pricing. Jeffrey MacKie-Mason, an Assoc iate Professor of Economics and Information, has designed the experiment at the University of Michigan. MacKie-Mason is also the Project Director for the economic aspects of the experiment.

This pricing field trial is called "Pricing Electronic Access to Knowledge" (PEAK). Michigan will create a variety of access models and administer a pricing system. The University will apply these models to other institutions, which will be serviced from Michigan acting as the host facility. Some will purchase access on a more or less standard subscription model. Others will buy a generalized or virtual subscription, which allows for prepaid access to a set of N articles, where the articles can be selected from across the database. Finally, a third group will acquire articles strictly on a transactional basis. Careful thought has, of course, gone into the relationship among the unit prices under these three schemes, the absolute level of the prices and the relationship between the pricing, concepts of value and the publishers' need for a return.

The experiment should begin in the summer of 1997 and run at least through 1998. We are all looking forward to the results of this research.

IN CONCLUSION

Journal publishers have relatively little experience with offering electronic full text to end users for a fee. Most new Web products either are free or have a free introductory period. Many are now in the process of starting to charge (*Science*, for example, instituted its first subscription fees as of January, 1997, and will only sell electronic subscriptions to paper personal subscribers). However, it is already clear that a price that is perceived as fair is a necessary but not sufficient factor in gaining users. Freely available information will not be used if it is not seen as being a productive use of time. Novelty fades quickly. If a Web site or other electronic offering does not offer more (job leads, competitive information, early reporting of research results, discussion forums, simple convenience of bringing key journals to the desktop), it will not be heavily used. In designing electronic services, publishers have to deal with issues of speed, quality control, comprehensiveness -- and then price. The evaluation of acceptance by the user will be on the total package.

* * *

For additional information about the conference, or The Andrew W. Mellon Foundation's scholarly communication initiatives, please contact Richard Ekman. For additional information about ARL or this web site contact Patricia Brennan, ARL Program Officer at (202) 296-2296.

Return to Office of Scholarly Communication Home Page

ARL Home ARL Scholarly Communication and Technology Home Page

© Association of Research Libraries, Washington, DC

Web Design by Angelo F. Cruz
Maintained by ARL Web Administrator
Last Modified: August 1, 1997

Scholarly Communication and Technology

Conference Organized by The Andrew W. Mellon Foundation

at Emory University
April 24-25, 1997

Copyright © of the papers on this site are held by the individual authors or The Andrew W. Mellon Foundation. Permission is granted to reproduce and distribute copies of these works for nonprofit educational or library purposes, provided that the author, source, and copyright notice are included on each copy. For commercial use, please contact Richard Ekman at the The Andrew W. Mellon Foundation.

Session #3 Economics of Electronic Publishing: Journals Pricing and User Acceptance

The Economics of Electronic Journals

Andrew Odlyzko
AT&T Labs - Research
amo@research.att.com

* * *

Revised draft, July 15, 1997.

Abstract

Can electronic publications be operated at much lower costs than print journals, and still provide all the services that scholars require? That is the key question that is still in dispute. Available evidence shows that free or at least much less expensive journals are possible on the Net. It is probable that such journals will dominate in the area of basic scholarly publishing. However, the transition is likely to be complicated, since the scholarly publishing business is full of inertia and perverse economic incentives.

1. Introduction

It is now practically universally accepted that scholarly journals will have to be available in digital formats. What is not settled is whether they can be much less expensive than print journals. Most traditional print publishers still claim, just as they have claimed for years, that switching to an electronic format can save at most 30% of the costs, namely the expenses of printing and mailing. Prices of electronic versions of established print journals are little, if any, lower than those of the basic paper versions. What publishers talk about most in connection with electronic publishing are the extra costs they bear, not savings [BoyceD]. On the other hand, there is also rapid growth of electronic-only journals run by scholars themselves, and available for free on the Internet.

Will the free electronic journals dominate? Most publishers claim that they will not survive (see, for example, [Babbitt]) and will be replaced by electronic subscription journals. Even some editors of the free journals agree with that assessment. My opinion is that it is too early to tell whether subscriptions will be required. It is likely that for we will have a mix of free and subscription journals, and that for an extended period neither will dominate. However, I am convinced that even the subscription journals will be much less expensive than the current print journals. The two main reasons are that modern technology makes it possible to provide the required services much more cheaply, and that in scholarly publishing, authors have no incentive to cooperate with the publishers in maintaining a high overhead system.

Section 2 summarizes the economics of the current print journal system. Section 3 looks at the electronic-only journals that have sprung up over the last few years and are available for free on the Net. Section 4 discusses the strange economic incentives that exist in scholarly publishing. Finally, Section 5 presents some tentative conclusions and projections.

This article draws heavily on my two previous papers on scholarly publishing, [Odlyzko1, Odlyzko2], and the references given there. For other references on electronic journals, see also [Bailey, PeekN]. It should be stressed that only scholarly journal publishing is addressed here. Trade publishing will also be revolutionized by new technology. However, institutional and economic incentives are different there, so the outcome will be different.

Scholarly publishing is a public good, paid for largely (although often indirectly) by taxpayers, student's parents, and donors. The basic assumption I am making in this article is that its costs should be minimized to the largest extent consistent with delivering the services that scholars and the society they serve require.

2. Costs of print journals

Just how expensive is the current print journal system? While careful studies of the entire scholarly journal system had been conducted in the 1970s [KingMR, Machlup], they were obsolete by the 1990s. Recent studies, such as those in [AMSS, Kirby], address primarily prices that libraries pay, and they show great disparities. For example, among the mathematics journals considered in [Kirby], the price per page ranged from \$0.07 to \$1.53, and the price per 10,000 characters, which compensates for different formats, from under \$0.30 to over \$3.00. Such statistics are of greatest value in selecting journals to purchase or (much more frequently) to drop, especially when combined with measures of the value of journals, such as the impact factors calculated by the Science Citation Index. However, they are not entirely adequate when studying the entire scholarly journal publishing system. For example, in the statistics of [Kirby], the Duke Mathematics Journal (DMJ), published by Duke University Press, is among the least expensive one, at $0.19 per page. On the other hand, using the same methodology as that in [Kirby], the International Mathematics Research Notices (IMRN), coming from the same publisher as DMJ, would have been among the most expensive ones several years ago, and would be around the median now (as its size has expanded, while the price stayed about constant). The difference appears to come from the much smaller circulation of IMRN than of DMJ, and not from any inefficiencies or profits at Duke University Press. (This case is considered in more detail in Section 4.)

To estimate the systems cost of the scholarly journal publishing system, it seems advisable to consider total costs associated with an article. In writing the "Tragic loss ..." essay [Odlyzko1], I made some estimates based on a sample of journals, all in mathematics and computer science. They were primary research journals, purchased mainly by libraries. The main identifiable costs associated with a typical article were the following:

1. revenue of publisher: $4,000
2. library costs other than purchase of journals and books: $8,000
3. editorial and refereeing costs: $4,000
4. authors' costs of preparing a paper: $20,000

Of these costs, the publishers' revenue of $4,000 per article (i.e., the total revenue from sales of a journal, divided by the number of articles published in that journal) is the one that attracts the most attention in discussions of the library or journal publishing "crises." It is also the one that is easiest to measure and most reliable. However, it is also among the smallest, and this is a key factor in the economics of scholarly publishing. The direct costs of a journal article are dwarfed by various indirect costs and subsidies.

The cost estimates above are only rough approximations, especially those for the indirect costs of preparing a paper. There is no accounting mechanism in place to associate the costs in items (3) and (4) with budgets of academic departments. However, those costs are there, and they are large, whether they are half or twice the estimates presented here.

Even the revenue estimate (1) is a rough approximation. Most publishers treat their revenue and circulation data as confidential. There are some detailed accounts, such as that for the Americal Physical Society (APS) publications in [Lustig], and for the Pacific Journal of Mathematics in [Kirby], but they are few.

The estimate of $4,000 in publishers' revenue per article made in [Odlyzko1] has until recently been just about the only one available in the literature. It is supported by the recent study of Tenopir and King [TenopirK], which also estimates that the total costs of preparing the first copy of an article are around $4,000. The estimate in [Odlyzko1] was based primarily on data in [AMSS], and so is about five years out of date. If I were redoing my study, I would adjust for the rapid inflation in journal prices in the intervening period, which would inflate the costs. On the other hand, in discussing general scholarly publishing, I would probably deflate my estimate to account for the shorter articles that are prevalent in most areas. (The various figures for size of the literature and so on derived in [Odlyzko1] were based on samples almost exclusively from mathematics and theoretical computer science, which were estimated to have articles of about 20 pages each. This is consistent with the data for these areas in [TenopirK]. However, the average length of an article over all areas is about 12 pages.) Thus, on balance, the final estimate for the entire scholarly literature would probably still be $3,000-4,000 as the publisher revenue from each article.

The $4,000 revenue figure was the median of an extremely dispersed sample. Among the journals used in [Odlyzko1] to derive that estimate, the cost per article ranged from under $1,000 for some journals to over $8,000 for others. This disparity in costs brings out another of the most important features of scholarly publishing, namely lack of price competition. Could any airline survive with $8,000 fares if a competitor offered $1,000 fares?

Wide variations in prices for seemingly similar goods are common even in competitive markets, but they are usually associated with substantial differences in quality. For example, one can sometimes purchase round-trip trans-Atlantic tickets for under $400, provided one travels in the off-season in coach, purchases them when the special sales are announced, travels on certain days, and so on. On the other hand, a first-class unrestricted ticket bought at the gate for the same plane can cost 10 times as much. However, it is easy to tell what the difference in price buys in this case. It is much harder to do so in scholarly publishing. There is some positive correlation between quality of presentation (proofreading, typography, and so on) and price, but it is not strong. In the area that matters the most to scholars, that of quality of material published, it is hard to discern any correlation. In mathematics, the three most prestigious journals are published by a commercial publisher, by a university, and by a professional society, respectively, at widely different costs. (Library subscription costs per page differ by more than a factor of 7 [Kirby], and it is unlikely that numbers of subscribers differ by that much.) In economics, the most prestigious journals are published by a professional society, the American Economic Association, and are among the least expensive ones in that field.

Many publishers argue that costs cannot be reduced much, even with electronic publishing, since most of the cost is the first-copy cost of preparing the manuscripts for publication. This argument is refuted by the widely differing costs among publishers. The great disparity in costs among journals is a sign of an industry that has not had to worry about efficiency. Another sign of lack of effective price competition is the existence of large profits. The economic function of high profits is to attract competition and innovation, which then reduce those profits to average levels. However, as an example, Elsevier's pretax margin exceeds 40% [Hayes], a level that is "phenomenally high, comparable as a fraction of revenues to the profits West Publishing derives from the Westlaw legal information service, and to those of Microsoft" [Odlyzko2]. Even professional societies earn substantial profits on their publishing operations.

> Not-for-profit scientific societies, particularly in the United States and in the UK, also often realize substantial surpluses from their publishing operations. ... Net returns of 30% and more have not been uncommon.
>
> [Lustig]

Such surpluses are used to support other activities of the societies, but in economic terms they are profits. Another sign of an industry with little effective competition is that some publishers keep over 75% of the revenues from journals just for distributing those journals, with all the work of editing and printing being done by learned societies.

While profits are often high in scholarly publishing, it is best to consider them just as an indicator of an inefficient market. While they are a substantial contributor to the journal crisis, they are not its primary cause. Recall that the publisher revenue of $4,000 per article is only half of the $8,000 library cost (i.e., costs of buildings, staff, and so on) associated with that article. Thus even if all publishers gave away their journals for free, there would still be a cost problem. The growth in the scholarly literature is the main culprit.

Even in the print medium, costs can be reduced. That they have not been is due to the strange economics of scholarly publishing, which will be discussed in Section 4. However, even the least expensive print publishers still operate at a cost of around $1,000 per article. Electronic publishing offers the possibility of going far below even that figure, as well as of dramatically lowering library costs.

3. Costs of "free" electronic journals

How low can the costs of electronic publishing be? One extreme example is provided by Paul Ginsparg's preprint server [Ginsparg]. It currently processes about 20,000 papers per year. These 20,000 papers would cost $40-80M to publish in conventional print journals (and most of them do get published in such journals, creating costs of $40-80M to society). To operate the Ginsparg server in its present state would take perhaps half the time of a systems administrator, plus depreciation and maintenance on the hardware (an ordinary workstation with what is by today's standards a modest disk system). This might come (with overheads) to a maximum of $100K per year, or about $5 per paper.

In presentations by publishers, one often hears allusions to big NSF grants and various hidden costs in Ginsparg's operation. Ginsparg does have a grant from NSF for $1M, spread over three years, but it is for software development, not for the operation of his server. However, let us take an extreme position, and let us suppose that he has an annual subsidy of $1M. Let us suppose that he spends all his time on the server (which he manifestly does not, as anyone who checks his publications record will realize), and let us toss in a figure of $300K for his pay (including the largest overhead one can imagine that even a high-overhead place like Los Alamos might have). Let us also assume that a new workstation had to be bought each year for the project, say at $20K, and let us multiply that by 5 to cover the costs of mirror sites. Let us in addition toss in $100K per year for several T1 lines just for this project. Even with all these outrageous overestimates, we can barely come to the vicinity of $1.5M per year, or $75 per paper. That is dramatically less than the $2,000-4,000 per paper that print journals require. (I am using a figure of $2,000 for each paper here as well as that of $4,000 from [Odlyzko1] since APS, the publisher of the lion's share of the papers in Ginsparg's server, and among the most efficient publishers, collects revenues of about $2,000 per paper.) As Andy Grove of Intel points out [Grove], any time anything important changes in a business by a factor of 10, it is necessary to rethink the whole enterprise. Ginsparg's server lowers costs by about two orders of magnitude, not just one.

A skeptic might point out that there are other "hidden subsidies" that have not been counted yet, such as those for the use of the Internet by the users of Ginsparg's server. Those costs are there, although the bulk of them is not for the Internet, which is comparatively inexpensive, but for the workstations, local area networks, and users' time coping with buggy operating systems. However, those costs would be there no matter how scholarly papers are published. Publishers depend on the postal system to function, yet are not charged the entire cost of that system.

Similarly, electronic publishing is a tiny part of the load on the computing and communications infrastructure, and so should not be allocated much of the total cost.

Ginsparg's server is an extreme example of minimizing costs. It also minimizes service. There is no filtering of submissions, nor any editing, the things that distinguish a journal from a preprint archive. Some argue that no filtering is necessary, and that preprints are sufficient to allow the community to function. However, such views are rare, and most scholars agree that journals do perform an important role. Even though some argue that print plays an essential role in the functioning of the journal system (see the arguments in [Rowland] and [Harnad] for opposing views on this issue), it appears that electronic journals can function just as well as print ones. The question in this paper is whether financial costs can be reduced by switching to electronic publishing.

There are hundreds of electronic journals that are operated by their editors and are available for free on the Net. They do provide all the filtering that their print counterparts do. However, although their ranks appear to double every year [ARL], they are all new and small. The question is whether a system of free journals is durable, and whether it can be scaled to cover most of scholarly publishing.

Two factors make free electronic journals possible. One is advances in technology, which make it possible for scholars to handle tasks such as typesetting and distribution that used to require trained experts and a large infrastructure. The other factor is a peculiarity of the scholarly journal system that has already been pointed out above. The monetary cost of the time that scholars put into the journal business as editors and referees is about as large as the total revenue that publishers derive from sales of the journals. Scholarly journal publishing could not exist in its present form if scholars were compensated financially for their work. Technology is making their tasks progressively easier. They could take on new roles and still end up devoting less effort to running the journal system.

Most scholars are already typesetting their own papers. Many were forced to do so by cutbacks in secretarial support. However, even among those, there are few who would go back to the old system of depending on technical typists if they had a choice. Technology is making it easier to do many tasks oneself than to explain to others how to do them.

Editors and referees are increasingly processing electronic submissions, even for journals that appear exclusively in print. Moreover, the general consensus is that this makes their life much easier. Therefore, if the additional load of publishing an electronic journal were small enough, one might expect scholars to do everything themselves. That is what many editors of the free electronic journals think is feasible. As the volume of papers increases, one can add more editors to spread the load, as the Electr. J. Comb. [EJC] has done recently (and as print journals have done in the past). The counterargument (cf. [Babbitt, BoyceD]) is that there will always be too many repetitive and tedious tasks to do, and that even those scholars who enjoy doing them now, while they are a novelty, will get tired of them in the long run. If so, it will be necessary to charge for access to electronic journals to pay for the expert help needed to run them. Some editors of the currently-free electronic journals share this view. However, none of the estimates of what would be required to produce acceptable quality come anywhere near the $4,000 per article that current print publishers collect. In [Odlyzko1] I estimated that $300-1,000 per article should suffice, and many others, such as Stevan Harnad, have come up with similar figures. In the years since [Odlyzko1] was written, much more experience in operations of free electronic-only journals has been acquired. I have corresponded and had discussions with editors of many journals, both traditional print-only, and free electronic-only. The range of estimates of what it would cost to run a journal without requiring authors, editors, and referees to do noticeably more than they are doing now is illustrated by the following two examples (both from editors of print-only journals):

a. The Editor-in-Chief of a large journal, which publishes around 200 papers per year (and processes several times that many submissions) and brings in revenues of about $1M per year to the publisher thinks he could run an electronic journal of equivalent quality with a subsidy of about $50K per year to pay for an assistant to handle correspondence and minor technical issues. He feels that author-supplied copies are usually adequate, and that the work of technical editors at the publisher does not contribute much to the scientific quality of the journal. If he is right, then $250 per paper is sufficient.

b. An editor of a much smaller journal thinks that extensive editing of manuscripts is required. In his journal, he does all the editing himself, and the resulting files are then sent directly to the printer, without any technical staff at the publisher being involved. He estimates that he spends between 30 minutes and an hour per page, and thinks that having somebody with his professional training and technical skills do the work results in substantially better result. If we assume a loaded salary of $100K per year (since such work could often be done by graduate students and junior postdocs looking for some extra earnings in their spare time), we have an estimate of $25-50 per page, or $250-1,000 per article, as the cost of running an electronic journal of comparable quality.

All the estimates fit in the range $300-$1,000 per article that was projected in [Odlyzko1], and do not come close to the $4,000 per article charged by traditional publishers. Why is there such a disparity in views on costs? It is not caused by a simple ignorance of what it takes to run a viable journal on the part of advocates of free or low-priced publications, since many of them are running successful operations. The disparity arises out of different views of what is necessary.

It has always been much easier to enlarge a design or add new features than to slim down. This has been noted in ship design [Pugh], cars, and airplanes, as well as in computers, where the mainframe builders were brought to the brink of ruin (and often beyond) before they learned from the PC industry. Established publishers are increasingly providing electronic versions of their journals, but usually only in addition to the print version. It is no surprise therefore that their costs are not decreasing. The approach of the free electronic journal pioneers has been different, namely to provide only what can be done with the resources available. They are helped by what are variously called the 80-20 or 70-30 rules (the last 20% of what is provided costs 80% of the total, etc.). By throwing out a few features, it is possible to lower costs dramatically. Even in the area of electronic publishing, the spectrum of choices is large. Eric Hellman, editor of "The MRS Internet Journal of Nitride Semiconductor Research" [MRS], which provides free access to all readers, but charges authors $275 for each published papers, commented (private communication) that with electronic publishing,

> $250/paper gets you 90% of the quality that $1000/paper gets you.

Electronics is making it much clearer than ever that there are many choices in terms of quality and price in publishing.

An example of large differences in costs is provided by projects that make archival information available digitally. Astrophysicists are in the process of digitizing about a million pages of journal articles (without doing optical character recognition, OCR, on the output) and are making them available for free on the Web. The scanning project (paid for by a grant from NASA) is carried out in the U.S., yet still costs only $0.18 per page in spite of the high wages. On the other hand, the costs of the JSTOR project, which was cited in [Odlyzko2] as paying about $0.20 per page for scanning, are more complicated. JSTOR pays a contractor around $0.40 per page for a combination of scanning, OCR, and human verification of the OCR output, with the work done in a less-developed country that has low wage costs. However, JSTOR's total costs are much higher, about $1-2 per page, since they rely on trained professionals in the U.S. to ensure they have complete runs of journals, that articles are properly classified, and so on. Since JSTOR aims to provide libraries with functionality similar to that of bound volumes, it is natural for it to strive for high quality. This raises costs, unfortunately.

It is important to realize how easy it is to raise costs. Even though lack of price competition in scholarly publishing has created unusually high profits [Hayes], most of the price that is paid for journals covers skilled labor. The difference in costs between the astrophysics and JSTOR projects is dramatic, but it does not come from any extravagance. Even at $2 per page, the average scholarly article would cost around $25 to process. At a loaded salary of $100K per year for a trained professional, that $25 corresponds to only half an hour of that person's time. Clearly one can boost the costs by doing more, and JSTOR must be frugal in the use of skilled labor.

Is the higher quality of the JSTOR project worth the extra cost? It is probably essential for JSTOR to succeed in its mission, which is to eliminate the huge print collections of back issues of journals. Personally I feel that JSTOR is a great project, the only one I am aware of in scholarly publishing that benefits all three parties, scholars, libraries, and publishers. Whether it will succeed is another question. It does cost more than just basic scanning, and it does require access restrictions. One can argue that the best course of action would be simply to scan the literature right away, while there are still low-wage countries that will do the work inexpensively. The costs of the manual work of cutting open volumes and feeding sheets into scanners is not likely to become much smaller. At $0.20 per page, the entire scholarly literature could probably be scanned for less than $200M. (By comparison, the world is paying several billion dollars per year just for one year of current journals, and the Harvard libraries alone cost around $60M per year to operate.) Once the material was scanned, it would be available in the future for OCR and addition of other enhancements.

The main conclusion to be drawn from the discussion in this section is that the monetary costs of scholarly publishing can indeed be lowered, even in print. Whether they will be is another question, one closely bound up with the strange economics of the publishing industry.

4. The perverse incentives in scholarly publishing

Competition drives the economy, but it often works in strange ways. A study done a few years ago (before managed care became a serious factor) compared hospital costs in mid-sized U. S. cities that had either one or two hospitals. An obvious guess might be that competition between hospitals would lead to lower costs in cities that had two hospitals. However, the results were just the opposite, with the two-hospital cities having substantially higher prices. This did not mean that basic economic laws did not apply. Competition was operating, but at a different level. Since it was doctors who in practice determined what hospital a patient went to, hospitals were competing for doctors by purchasing more equipment, putting in specialty wards, and the like, which was increasing their costs (but not making any noticeable difference in the health of the population they served). The patients (or, more precisely, their insurers and employers) were paying the extra price.

Scholarly publishing as a business has many similarities to the medical system, except that if anything, it is even more involved. Journals do not compete on price, since that is not what determines their success. There are four principal groups of players. The first one consists of scholars as producers of the information that makes journals valuable. The second consists of scholars as users of that information. However, as users, they gain access to journals primarily through the third group, the libraries. Libraries purchase journals from the fourth group, the publishers, usually in response to requests from scholars. These requests are based overwhelmingly on the perceived quality of the journals, and price seldom plays a role (although that is changing under the pressure to control growth of library costs). The budgets for libraries almost always come from different sources than the budgets for academic departments, so that scholars as users do not have to make an explicit tradeoff between graduate assistantships and libraries, for example.

Scholars as writers of papers determine what journals their work will appear in, and thus how much it will cost society to publish their work. However, scholars have no incentive to care about those costs. What matters the most to them is the prestige of the journals they publish in. Often the economic incentives are to publish in high-cost outlets. It has often been argued that page charges are a rational way to allocate costs of publishing, since they make the author (or the author's institution or research grant) cover some of the costs of the publication, which, after all, is motivated by a desire to further the author's career. However, page charges are less and less frequent. As an extreme example, in the late 1970s, Nuclear Physics B, published by Elsevier, took over as the "journal of choice" in particle physics and field theory from Physical Review D, even though the latter was much less expensive. This happened because Phys. Rev. D had page charges, and physicists decided they would rather use their grant money for travel, postdocs, and the like. Note that the physicists in this story behaved in a perfectly rational way. They did not have to use their grants to pay for the increase in library costs associated with the shift from an inexpensive journal to a much pricier one. Furthermore, even if they had to pay for that cost, they would have come out ahead; the increase in the costs of just their own library associated with an individual decision to publish in Nucl. Phys. B instead of the less expensive Phys. Rev. D (could such a small change have been quantified) would have been much smaller than the savings on page charges. Most of the extra cost would have been absorbed by other institutions.

To make this argument more explicit, consider two journals, H (high priced) and L (low priced). Suppose that each one has 1,000 library subscriptions and no individual ones. L is a lean operation, and it costs them $3,000 to publish each article. They collect $1,000 from authors through page charges, and the other $2,000 from subscribers, so that each library in effect pays $2 for each article that appears in L. On the other hand, H collects $7,000 in revenue per article, all from subscriptions, which comes to $7 per article for each library. (It does not matter much whether the extra cost of H is due to profits, higher quality, or inefficiency.)

From the standpoint of the research enterprise, or of any individual library, it would be desirable to steer all authors towards publishing in L, as that would save a total of $4,000 for each article. However, look at this situation from the standpoint of the author. If she publishes in L, she loses

$1,000 that could be spent on graduate students, conferences, etc. If she publishes in H, she gets to keep that money. She does not get charged for the extra cost to any library, at least not right away. Eventually the overhead rates on her contract might go up to pay for the higher library spending at her institution. However, this effect is delayed and is weak. Even if we had accounting mechanisms that would provide instantaneous feedback (which we do not, with journal prices set over a year in advance and totally insensitive to minor changes caused by individual authors deciding where to publish), our hypothetical author would surely only get charged for the extra $5 that she causes her library to spend ($7 for publication in H as opposed to $2 in L), and not for the costs to all the other 999 libraries. She would still save $995 ($1000 - $5) of her grant money. Is it any wonder if she chooses to publish in H?

A secondary consideration for authors is to ensure that their papers are widely available. However, this factor has seldom played a major role, and with the availability of preprints through email or home pages it is becoming even less significant. Authors are not told what the circulation of a journal is (although for established publications, they probably have a rough idea of how easy it is to access them). Further, it is doubtful this information would make much difference, at least in most areas. Authors can alert the audience they really care about (typically a few dozen experts) through preprints, and the journal publication is for the resume more than to contact readers.

In 1993-4, there was a big flap about the pricing of International Mathematics Research Notices (IMRN), a new research announcement journal spun off from the Duke Mathematical Journal. The institutional subscriptions cost $600 per year, and there were not many papers in it. The Director of Publishing Operations for Duke University Press then responded in the Newsletter on Serials Pricing Issues [NSPI], by saying that his press was doing the best it could to hold down prices. It's just that their costs for IMRN were going to be $60,000 per year, and they expected to have 100 (sic!) subscriptions, so they felt they had to charge $600 per subscription. Now one possibility is that the Duke University Press miscalculated, and that it might have been easier for them to sell 400 subscriptions at $150 than 100 at $600, since IMRN did establish a good reputation as an insert to Duke Math. J. However, if their decision was right, then there seem to be two possibilities: (i) scholars will decide that it does not make sense to publish in a journal that is available in only 100 libraries around the world, or (ii) scholars will continue submitting their papers to the most prestigious journals they can find (such as IMRN), no matter how small their circulation, since prestige is what counts in tenure and promotion decisions, and since everybody that they want to read their papers will be able to get them electronically from preprint servers in any case. In neither case are journals such as IMRN likely to survive in their present form. (IMRN itself appears to have gained a longer lease on life, since is seems to have gained considerably more subscribers, and while it has not lowered its price, it is publishing many more papers, lowering its price per page, as mentioned in Section 2.)

The perverse incentives in scholarly publishing that are illustrated in the examples above have led to the current expensive system. They are also leading to its collapse. The central problem is that scholars have no incentive to maintain it. In book publishing, royalties align the authors' interests with those of publishers, as both wish to maximize revenues. (This is most applicable in the trade press, or in textbooks. In scholarly monograph publishing, the decreasing sales combined with the typical royalty rate of at most 15% are reducing the financial payoff to authors, and appears to be leading to changes, with monographs becoming available electronically for free.) For the bulk of scholarly publishing, though, the market is too small to make provide a significant financial payoff to the authors.

5. The future

Although scholars have no incentive to maintain the current journal system, they currently also have no incentive to dismantle it. Even the physicists who rely on the Ginsparg preprint server continue to publish most of their papers in established print journals. The reason is that it costs them nothing to submit papers to such journals, and also costs them nothing to have their library buy the journals. The data from the Association of Research Libraries [ARL] show that the average cost of the library system at leading research universities is about $12,000 per faculty member. (It is far higher at some, with Princeton spending about $30,000 per year per faculty member.) This figure, however, is not visible to the scholars, and they have no control over it. They are not given a choice between spending for the library and for other purposes.

Until the academic library system is modified, with the costs and tradeoffs made clear to scholars and administrators, it is unlikely there will be any drastic changes. We are likely to see slow evolution (cf. [Odlyzko3]), with continuing spread of preprints (in spite of attempts of journals in certain areas, such as medicine, to play King Canute roles, and attempt to stem this natural growth). Electronic journals will become almost universal but most of them will be versions of established print journals, and will be equally expensive. Free or inexpensive electronic journals will grow, but probably not too rapidly. However, this situation is not likely to persist for long. I have been predicting [Odlyzko1, Odlyzko2] that change will come when administrators realize just how expensive the library system is, and that scholars can obtain most of the information they need from other sources, primarily preprints. Over the decade from 1982 to 1992, library expenditures have grown by over a third even after adjusting for general inflation [ARL]. However, they have fallen by about 10% as a share of total university spending. Apparently the pressure from scholars to maintain library collection has not been great enough, and other priorities have been winning. At some point in the future more drastic cuts are likely.

How cuts will be distributed is unclear. We are entering the Information Age, and total spending on information is unlikely to decrease, but it probably will move into new channels. In discussions of the library crisis, most attention is devoted to journal costs. However, for each $1 spent on journal acquisitions, other library costs come to $2. If publishers can provide electronic versions of not only their current issues, but also older ones (either themselves or through JSTOR), they can improve access to scholarly materials and lower the costs of the library system (buildings, staff, maintenance) without lowering their own revenues. It is doubtful whether that will be enough, though, and it is likely that spending on journals as well as the rest of the library system will decrease.

Acknowledgements. I thank Erik Brynjolfsson, Joe Buhler, Peter Denning, Mark Doyle, Paul Ginsparg, Stevan Harnad, Steve Heller, Eric Hellman, Don King, Rob Kirby, Gene Klotz, Silvio Levy, Laszlo Lovasz, Harry Lustig, Robert Miner, Ann Okerson, Bernard Rous, Arthur Smith, Ron Stern, Edward Vielmetti, Lars Wahlbin, Bernd Wegner, and Ronald Wigington for their comments and the information they provided.

REFERENCES:

[AMSS] Survey of American research journals, Notices Amer. Math. Soc. 40 (1993), 1339-1344.

[ARL] Association of Research Libraries, </>.

[Babbitt] D. G. Babbitt, Mathematical journals: Past, present and future - a personal view, Notices Amer. Math. Soc., Jan. 1997. Available at <http://www.ams.org/>.

[Bailey] C. W. Bailey, Jr., Scholarly electronic publishing bibliography, available at <http://info.lib.uh.edu/sepb/sepb.html> .

[BoyceD] P. B. Boyce and H. Dalterio, Electronic publishing of scientific journals, Physics Today, Jan. 1996, pp. 42-47. Available at <http://www.aas.org/~pboyce/epubs/> .

[EJC] The Electronic Journal of Combinatorics, <http://www.combinatorics.org/> .

[EMIS] European Mathematical Information Service, <http://www.emis.de/> .

[Ginsparg] P. Ginsparg, Winners and losers in the global research village, available at <http://xxx.lanl.gov/blurb/pg96unesco.html> .

[Grove] A. Grove, "Only the Paranoid Survive," Bantam Doubleday Dell, 1996.

[Harnad] S. Harnad, The paper house of cards (and why it's taking so long to collapse), Ariadne, issue # 8, March 1997. Web version at <http://www.ukoln.ac.uk/ariadne/> .

[Hayes] J. R. Hayes, The Internet's first victim?, Forbes, Dec. 18, 1995, pp. 200-201.

[KingMR] D. W. King, D. D. McDonald, and N. K. Roderer, "Scientific Journals in the United State. Their production, Use and Economics," Hutchinson Ross, 1981.

[Kirby] R. Kirby, Comparative prices of math journals, available at <http://math.berkeley.edu/~kirby/journals.html> .

[Lustig] H. Lustig, Electronic publishing: economic issues in a time of transition, in "Electronic Publishing for Physics and Astronomy," A. Heck, ed., Kluwer, 1997, to appear.

[Machlup] F. Machlup, K. Leeson, and Associates, "Information Through the Printed Word: The Dissemination of Scholarly, Scientific, and Intellectual Knowledge," vol. 2: Journals, Praeger, 1978.

[MRS] The MRS Internet Journal of Nitride Semiconductor Research, available at <http://nsr.mij.mrs.org/> .

[NSPI] Newsletter on Serials Pricing Issues, published electronically at Univ. of North Carolina, available at URL <http://sunsite.unc.edu/reference/prices/prices.html> .

[Odlyzko1] A. M. Odlyzko, Tragic loss or good riddance? The impending demise of traditional scholarly journals, Intern. J. Human-Computer Studies (formerly Intern. J. Man-Machine

Studies) 42 (1995), 71-122. Also in the electronic J. Univ. Comp. Sci., pilot issue, 1994, <http://hyperg.iicm.tu-graz.ac.at/> . Available at <http://www.research.att.com/~amo/> .

[Odlyzko2] A. M. Odlyzko, On the road to electronic publishing, Euromath Bulletin, vol. 2, no. 1 (June 1996), 49-60. Available at author's home page, <http://www.research.att.com/~amo/> .

[Odlyzko3] A. M. Odlyzko, The slow evolution of electronic publishing. In preparation. Will be available at <http://www.research.att.com/~amo/> .

[PeekN] R. P. Peek and G. B. Newby, eds., "Scholarly Publishing: The Electronic Frontier," MIT Press, 1996.

[Pugh] P. Pugh, "The Cost of Seapower: The Influence of Money on Naval Affairs from 1815 to the Present Day," Conway, 1986.

[Rowland] F. Rowland, Print journals: Fit for the future?, Ariadne, issue # 7, Jan. 1997. Web version at <http://www.ukoln.ac.uk/ariadne/> .

[TenopirK] C. Tenopir and D. W. King, Trends in scientific scholarly journal publishing in the U. S., J. Scholarly Publishing, April 1997, to appear.

……………………………………

For additional information about the conference, or The Andrew W. Mellon Foundation's scholarly communication initiatives, please contact Richard Ekman. For additional information about ARL or this web site contact Patricia Brennan, ARL Program Officer at (202) 296-2296.

Return to Office of Scholarly Communication Home Page

ARL Home ARL Scholarly Communication and Technology Home Page

© Association of Research Libraries, Washington, DC
Web Design by Angelo F. Cruz
Maintained by ARL Web Administrator
Last Modified: August 14, 1997

Scholarly Communication and Technology

Conference Organized by The Andrew W. Mellon Foundation

at Emory University
April 24-25, 1997

Copyright © of the papers on this site are held by the individual authors or The Andrew W. Mellon Foundation. Permission is granted to reproduce and distribute copies of these works for nonprofit educational or library purposes, provided that the author, source, and copyright notice are included on each copy. For commercial use, please contact Richard Ekman at the The Andrew W. Mellon Foundation.

Session #4 Patterns of Usage

Collaboratory for Research on Electronic Work

Analysis of JSTOR: The impact on scholarly practice of access to on-line journal archives

Thomas A. Finholt
Assistant Professor of Psychology
The University of Michigan

and

JoAnn M. Brooks
The University of Michigan

* * *

Running head: Analysis of JSTOR

Analysis of JSTOR: The impact on scholarly practice of access to on-line journal archives

Preparation of this article was supported by a grant to the University of Michigan from the Andrew W. Mellon Foundation. JSTOR ^{TM} is the proprietary product of JSTOR, Inc. a non-profit corporation dedicated to provision of digital access to the back archives of scholarly journals. For more information, please consult www.jstor.org/.

We gratefully acknowledge the assistance of Kristin Garlock, Marcia Heringa, Christina Maresca, William Mott, Sherry Piontek, Tony Ratanproeksa, Blake Sloan, and Melissa Stucki in gathering the data for this study. Also, we thank Ann Bishop, Joan Durrance, Kristin Garlock, Kevin Guthrie, Wendy Lougee, Sherry Piontek, and the participants of the Andrew W. Mellon Foundation Scholarly Communication and Technology Conference for comments on earlier drafts. Finally, we thank the history and economics faculty of Bryn Mawr College, Denison University, Haverford College, Swarthmore College, the University of Michigan, and Williams College for their patience and cooperation as participants in this research.

Requests for copies should be sent to: a) Thomas Finholt, Collaboratory for Research on Electronic Work, C-2420 701 Tappan Street, Ann Arbor, MI 48109-1234; or b) finholt@umich.edu

Abstract

This study reports on faculty response to the Journal STORage project (JSTOR), an on-line system for accessing digital back archives of core journals in history and economics. Data were collected about general journal use, Internet use, and JSTOR use via a survey administered to 160 historians and economists at the University of Michigan and at five liberal arts colleges: Bryn Mawr College, Denison University, Haverford College, Swarthmore College, and Williams College. Results show that most faculty do not yet use JSTOR. When JSTOR use occurs, frequency of use is positively related to being male, having a preference for photocopying journal articles, relying on article abstracts when reading journals, and the frequency of searching on-line card catalogs. Increased numbers of journal subscriptions and affiliation with an economics department are negatively related to the frequency of JSTOR use. The findings suggest that faculty may be willing to substitute access to digital journal back archives for access to bound journals, but this willingness may vary by discipline.

Analysis of JSTOR: The impact on scholarly practice of access to on-line journal archives

Innovations introduced over the last thirty years, such as computerized library catalogs and on-line citation indexes, have transformed scholarly practice. Today, the dramatic growth of worldwide computer networks raises the possibility for further changes in how scholars work. For example, attention has focused on the Internet as an unprecedented mechanism for expanding access to scholarly documents through electronic journals (Olsen, 1994; Odlyzko, 1995), digital libraries (Fox, Akscyn, Furuta, & Legett, 1995), and archives of pre-publication reports (Taubes, 1993). Unfortunately, the rapid evolution of the Internet makes it difficult to accurately predict which of the many experiments in digital provision of scholarly content will succeed. As an illustration, electronic journals have received only modest acceptance by scholars (Kling & Covi, 1996). Accurate assessment of the scholarly impact of the Internet requires attention to experiments that combine a high probability of success with the capacity for quick dissemination. According to these criteria, digital journal archives deserve further examination. A digital journal archive provides on-line access to the entire digitized back archive of a paper journal. Traditionally, scholars make heavy use of journal back archives in the form of bound periodicals. Therefore, providing back archive content on-line may significantly enhance access to a resource already in high demand. Further, studying the use of experimental digital journal archives may offer important insight into the design and functionality of a critical Internet-based research tool. This paper, then, reports on the experience of social scientists using the Journal STORage system. (JSTOR^{TM}), a prototype World Wide Web application for viewing and printing the back archives of ten core journals in history and economics.

The JSTOR system

JSTOR represents an experiment in the technology, politics, and economics of on-line provision of journal content. The technology involves scanning pages of paper journals to make bitmaps of these pages available for printing or for viewing on screen. In addition to the bitmaps, a text representation of each page exists. Search engines use the text representation to index the bitmaps of scanned pages, which then supports logical queries on the title, author, or full text of articles in the JSTOR system. JSTOR has a Web-based interface. This means that any user with access permission and a Web browser (e.g., Microsoft Internet Explorer) may search JSTOR. Through the same interface, users may view retrieved content -- exactly as it would appear in the paper journal -- and, via a helper application, users may print content. The JSTOR system can be previewed at http://www.jstor.org/.

The politics and economics of JSTOR involve complex issues of providing journal content to scholars without cannibalizing the market for paper journals. Specifically, journal publishing offers a lucrative source of revenue for private firms and for professional societies. To protect this revenue, JSTOR contains no current journal content. JSTOR does contain the entire back archive, within two to three years of the present, of core journals in a variety of disciplines. These back archives have tremendous value to scholars, but historically have not interested journal publishers due to the high cost of converting paper formats into digital formats. JSTOR attempts to price access to these back archives at a level conducive to universities and colleges, that is, below the carrying costs for handling and storing bound journals. The JSTOR mission, then, involves offering a service attractive to scholars, priced at a level acceptable to university and college libraries, and with sufficient revenue to ensure expansion and improvement of the JSTOR technology.

The initial rollout of JSTOR has involved librarians and faculty on six campuses. The current faculty audience for JSTOR consists of economists, historians, and ecologists -- reflecting the present content of JSTOR. This paper focuses on historians and economists using JSTOR at five private liberal arts colleges (Bryn Mawr College, Denison University, Haverford College, Swarthmore College, and Williams College) and one public research university (the University of Michigan). The core economics journals in JSTOR at the time of this study included: *American Economic Review*, *Econometrica*, *Quarterly Journal of Economics*, *Journal of Political Economy*, and *Review of Economics and Statistics*. The core history journals included: *American Historical Review*, *Journal of American History*, *Journal of Modern History*, *William and Mary Quarterly*, and *Speculum*. In the future, JSTOR will expand to include over 150 journal titles covering dozens of disciplines.

Journal use in the social sciences

To understand JSTOR use requires a general sense of how social scientists seek and use scholarly information. In practice, social scientists apply five main search strategies. First, social scientists use library catalogs. Broadbent (1986) found that 69% of a sample of historians used a card catalog when seeking information, while Lougee, Sandler, and Parker (1990) found that 97% of a sample of social scientists used a card catalog. Second, journal articles are a primary mechanism for communication among social scientists (Garvey, 1979; Garvey, Lin, & Nelson, 1970). For example, in a study of social science faculty at a large state university, Stenstrom and McBride (1979) found that a majority of the social scientists used citations in articles to locate information. Third, social scientists use indexes and specialty publications to locate information. As an illustration, Stenstrom and McBride (1979) found that 55% of social scientists in their sample reported at least occasional use of subject bibliographies and 50% reported at least occasional use of abstracting journals. Similarly, Olsen (1994) found that in a sample of sociologists 37.5% reported regular use of annual reviews. Fourth, social scientists browse library shelves. For instance, Lougee, et al. (1990) and Broadbent (1986) both found that social scientists preferred to locate materials by browsing shelves. Sabine and Sabine (1986) found that 20% of a sample of faculty library users reported locating their most recently accessed journal via browsing. On a related note, Stenstrom and McBride (1979) found that social scientists used departmental libraries more heavily than the general university library. Finally, social scientists rely on the advice of colleagues and students. For example, various studies show that colleagues have particular value when searching for a specific piece of information (Stenstrom & McBride 1979, Broadbent 1986, Simpson 1988). Also, students working on research projects often locate background material that social scientists find useful (Olsen, 1994; Simpson, 1988). Similarly, faculty report a valuable, but infrequent role for librarians in seeking information (Stenstrom & McBride, 1979; Broadbent, 1986; Lougee et al. 1990).

Computer-based tools do not figure prominently in the preceding description of how social scientists search for scholarly information. Results from previous studies show that the primary application of digital information technology for social scientists consists of computerized searching, which social scientists do at lower rates than physical scientists, but at higher rates than humanists (Lougee, et al. 1990; Olsen, 1994; Broadbent, 1986). Lougee, et al. (1990) and Olsen (1994) both report sparse use of on-line catalogs by social scientists. Evidence of the impact of demographic characteristics on use of digital resources is mixed. For example, Lougee, et al. (1990) found a negative correlation between age and use of digital information technology, while Stenstrom and McBride (1979) found no correlation. Finally, in a comparison of e-mail use by social scientists and humanists, Olsen (1994) found higher use rates among the social scientists, apparently correlated with superior access to technology.

In terms of journal access, previous studies indicate that economics faculty tend to subscribe to more journals than faculty in other social science disciplines (Simpson, 1988; Schuegraf & van Bommel, 1994). Journal subscriptions are often associated with membership in a professional society. For example, in their analysis of a liberal arts faculty, Schuegraf and van Bommel (1994) found that 40.9% of faculty journal subscriptions -- including 12 of the 15 most frequently subscribed journals -- came with society memberships. Stenstrom and McBride (1979) found that membership-related subscriptions often overlapped with library holdings. However, according to Schuegraf and van Bommel, other personal subscriptions included journals not held in library collections. In terms of journal use, Sabine and Sabine (1986) found that only 4% of faculty in their sample reported reading the entire contents of journals, while 9% reported reading single articles, and 87% reported reading only small parts, such as abstracts. Similarly, at least among a sample of sociologists, Olsen (1994) found that all respondents reported using abstracts to determine whether to read an article. Having found a relevant article, faculty often make copies. For instance, Sabine and Sabine (1986) found that

47% of their respondents had photocopied their most recently read journal article, Simpson (1988) found that 60% of sampled faculty reported "always" making copies, and all of the sociologists in Olsen's (1994) sample reported copying important articles.

Goals of this study

The research described above consists of work conducted prior to the advent of the World Wide Web and widespread access to the Internet. Several recent studies suggest that Internet use can change scholarly practice (Finholt & Olson, 1997; Hesse, Sproull, & Kiesler, 1994; Walsh & Bayma, 1997; Carley & Wendt, 1991). However, most of these studies focused on physical scientists. A key goal of this study is to create a snapshot of the effect of Internet use on social scientists, specifically use of JSTOR. Therefore, the sections that follow will address core questions about the behavior of JSTOR users, including: a) how faculty searched for information; b) which faculty used JSTOR; c) how journals were used d) how the Internet was used; and e) how journal use and Internet use correlated with JSTOR use.

Method

Participants

The population for this study consisted of the history and economics faculty at the University of Michigan and at five liberal arts colleges: Bryn Mawr College, Denison University, Haverford College, Swarthmore College, and Williams College. History and economics faculty were targeted because the initial JSTOR selections drew on ten journals, reflecting five core journals in each of these disciplines. The institutions were selected based on their status as Andrew W. Mellon Foundation grant recipients for the JSTOR project.

Potential respondents were identified from the roster of full-time history and economics faculty at each institution. With the permission of the respective department chairs at each school, faculty were invited to participate in the JSTOR study by completing a questionnaire. No incentives were offered for respondents and participation was voluntary. Respondents were told that answers would be confidential, but not anonymous due to plans for matching responses longitudinally. The resulting sample contained 161 respondents representing a response rate of 61%. In this sample, 46% of the respondents were economists, 76% were male, and 48% worked at the University of Michigan. The average respondent was 47.4 years old and had a Ph.D. granted in 1979.

Design and procedure

Respondents completed a 52 item questionnaire with questions on journal use, computer use, attitudes toward computing, information search behavior, demographic characteristics, and JSTOR use. Respondents had the choice of completing this questionnaire via a telephone interview, via the Web, or via a hardcopy version. Questionnaires were administered to faculty at the five liberal arts college and to the faculty at the University of Michigan in the spring of 1996.

Journal use. Journal use was assessed in four ways. First, respondents reported how they traditionally accessed the journal titles held in JSTOR, choosing from: no use; at the library; through a paid subscription; or through a subscription received with membership in a professional society. Second, respondents ranked the journals they used in order of frequency of use for a maximum of ten journals. For each of these journals, respondents indicated whether they had a personal subscription to the journal. Third, respondents described their general use of journals in terms of the frequency of browsing journal contents, photocopying journal contents, saving journal contents, putting journal contents on reserve, or passing journal contents along to colleagues (measured on a five point scale, where 1 = never, 2 = rarely, 3 = sometimes, 4 = frequently, and 5 = always). Finally, respondents indicated the sections of journals they used, including the table of contents, article abstracts, articles, book reviews, reference lists, and editorials.

Computer use. Computer use was assessed in three ways. First, respondents described their computer systems in terms of the type of computer (laptop vs. desktop), the computer family (e.g., Apple vs. DOS), the specific model (e.g., PowerPC), and the operating system (e.g., Windows95). Second, respondents reported their level of use via a direct network connection (e.g., Ethernet) of the World Wide Web, e-mail, databases, on-line library catalogs, and ftp (measured on a five point scale, where 1 = never, 2 = 2-3 times per year, 3 = monthly, 4 = weekly, and 5 = daily). Finally, respondents reported their level of use via a modem connection of the Web, email, databases, on-line library catalogs, and ftp (using the same scale as above).

Attitudes toward computing. Attitudes toward computing were assessed by respondents' reported level of agreement with statements about personal computer literacy, computer literacy relative to others, interest in computers, the importance of computers, confusion experienced while using computers, and the importance of programming knowledge (measured on a five point scale, where 1 = strongly disagree, 2 = disagree, 3 = neutral, 4 = agree, and 5 = strongly agree).

Information search behavior. Information search behavior was assessed in three ways. First, respondents indicated their use of general search strategies, including: searching/browsing on-line library catalogs; searching/browsing paper library catalogs; browsing library shelves; searching/browsing on-line indexes; searching/browsing paper indexes; browsing departmental collections; reading citations from articles; and consulting colleagues. Second, respondents described the frequency of literature searches within their own field and the frequency of on-line literature searches within their own field (both measured on a five point scale, where 1 = never, 2 = 2-3 times per year, 3 = monthly, 4 = weekly, and 5 = daily). Finally, respondents described the frequency of literature searches outside their field and the frequency of on-line literature searches outside their field (measured on the same five point scale used above).

Demographic characteristics. Respondents were asked to provide information on demographic characteristics, including: age, sex, disciplinary affiliation, institutional affiliation, highest degree attained, and year of highest degree.

JSTOR use. Finally, JSTOR use was assessed in two ways. First, respondents reported whether they had access to JSTOR. Second, respondents described the frequency of JSTOR use (measured on a five point scale, where 1 = never, 2 = 2-3 times per year, 3 = monthly, 4 = weekly, and 5 = daily).

Results

The data were analyzed to address five core questions related to the impact of JSTOR: a) how faculty searched for information; b) which faculty used JSTOR; c) how journals were used d) how the Internet was used; and e) how journal use and Internet use correlated with JSTOR use.

Information searching

Table 1 summarizes data on how faculty searched for information. Using citations from related publications (94%), consulting colleagues (90%), searching electronic catalogs (86%), browsing shelves (71%), browsing electronic catalogs (65%), using electronic indexes (64%), and using printed indexes (56%) were all strategies used by a majority of the faculty. A minority of the faculty reported using paper card catalogs (26%), browsing departmental collections (22%), and browsing paper card catalogs (16%). The proportion of faculty using the search strategies did not differ significantly by institution or discipline, with the exception of three strategies. First, the proportion of Michigan economists who reported browsing library shelves (46%) was significantly less than the proportion of five college historians who used this strategy (86%). Second, the proportion of Michigan economists who reported searching card catalogs (14%) was significantly less than the proportion of five college historians who used this strategy (39%). And finally, the proportion of Michigan economists who reported browsing departmental collections (48%) was significantly greater than the proportion of five college historians who used this strategy (4%)[1].

Who used JSTOR

Overall, 67% of the faculty did not use JSTOR[2], 14% used JSTOR once a year, 11% used JSTOR once a month, and 8% used JSTOR once a week. None of the faculty used JSTOR daily. Table 2 summarizes JSTOR frequency of use by type of institution and discipline. A comparison of use by type of institution shows a higher proportion of JSTOR users at the five colleges (42%) than at the University of Michigan (27%). A further breakdown by discipline shows that the five college economists had the highest proportion of users (46%), followed by the Michigan economists (40%), the five college historians (39%), and the Michigan historians (16%). One way to put JSTOR use into perspective is to compare this activity with similar, more familiar on-line activities, like literature searching. Overall, 21% of the faculty did not do on-line searches, 25% searched once a year, 25% searched once a month, 25% searched once a week, and 4% searched daily. Table 3 summarizes data on the frequency of on-line searching by type of institution and discipline for the same faculty described in Table 2. A comparison of on-line searching by type of institution shows a higher proportion of on-line searchers at the five colleges (85%) than at the University of Michigan (76%). A further breakdown by discipline shows that five college economists had the highest proportion of searchers (89%), followed by the five college historians (82%), and the Michigan economists and historians (both 76%).

Figure 1 shows a plot of the cumulative percentage of faculty per institution who used JSTOR and who did on-line searches versus the frequency of these activities. For example, looking at the values plotted on the y-axis against the "Monthly" category shows that over three times as many Michigan faculty searched once a month or more (51%) compared to the percentage of faculty who used JSTOR once a month or more (15%). Similarly, over two times as many of the five college faculty searched once a month or more (62%) compared to the percentage of faculty who used JSTOR once a month or more (25%). A further breakdown by discipline shows that over twice as many of the five college economists searched once a month or more (73%) compared to once a month or more use of JSTOR (31%), that over six times as many of the Michigan historians searched once a month or more (54%) compared to once a month or more use of JSTOR (8%), that over twice as many of the five college historians searched once a month or more (50%) compared to once a month or more use of JSTOR (21%), and that over twice as many of the Michigan economists searched once a month or more (48%) compared to once a month or more use of JSTOR (23%).

Journal use

Table 4 summarizes how faculty used features of journals. Articles were the most used feature (used by 98% of the faculty) and editorials were the least used feature (used by 26% of the faculty). Across all journal features, patterns of use were similar, except in two areas. First, the proportion of Michigan historians who used article abstracts (31%) was significantly smaller than the proportion of Michigan economists (81%), five college economists (89%), and five college historians (61%) who used abstracts. Second, the proportion of Michigan economists who used book reviews (49%) was significantly smaller than the proportion of five college historians (100%), Michigan historians (98%), and five college economists (85%) who used book reviews.

Overall, faculty in the sample reported that they regularly used 8.7 journals, that they subscribed to 4.1 of these journals, and that 2.2 of these journals were also in JSTOR. Table 5 summarizes journal use by institution and discipline. There were no significant differences in the number of journals used across institution and discipline, although Michigan historians reported using the most journals (8.9). There were also no significant differences across institution and discipline in the number of paid journal subscriptions among the journals used, although again Michigan historians reported having the most paid subscriptions (4.6). There was a significant difference in the number of journals used regularly by the economists that were also titles in JSTOR ($M = 2.9$), compared to the historians ($M = 1.7$), $t(158) = 5.71$, $p<.01$.

Further examination of differences in use of journals shows a much greater consensus among the economists about the importance of the economics journals in JSTOR than among the historians about the history journals in JSTOR. For example, Table 6 shows the economists' ranking in order of use of the five economics journals chosen for JSTOR. The *American Economic Review* was cited among the top ten most frequently used journals by over 75% of both the Michigan and the five college economists, the *Journal of Political Economy* was cited among the top ten by over 60% of both the Michigan and the five college economists, and the *Quarterly Journal of Economics* and the *Review of Economics and Statistics* were cited among the top ten by over 50% of the Michigan economists and by over 40% of the five college economists. By contrast, Table 7 shows the historians' ranking in order of use of the five history journals chosen for JSTOR. The *American Historical Review* was cited among the top ten most frequently used journals by over 60% of both the Michigan and the five college historians. However, none of the other four journals were used by a majority of the historians at Michigan or at the five colleges.

Internet use

Overall, faculty reported weekly use of email (*M* = 4.3), monthly use of on-line catalogs (*M* = 3.2) and the Web (*M* = 3.0), and two or three uses per year of ftp (*M* = 2.3) and on-line database (*M* = 2.1). Table 8 summarizes the use of these Internet applications by institution and discipline. In terms of email use, Michigan historians (*M* = 3.3) were significantly lower than the Michigan economists (*M* = 4.9), the five college economists (*M* = 5.0), and the five college historians (*M* = 4.7). In terms of World Wide Web use, Michigan historians (*M* = 1.8) were significantly lower than everyone, while the five college historians (M = 2.9) were significantly lower than the five college economists (*M* = 4.2) and the Michigan economists (*M* = 3.9). In terms of ftp use, the Michigan historians (*M* = 1.4) and the five college historians (*M* = 1.7) differed significantly from the Michigan economists (*M* = 3.4) and the five college economists (*M* = 2.7). In terms of on-line database use, the Michigan historians (*M* = 1.6) were significantly lower than the five college economists (*M* = 2.9). Faculty did not differ significantly in terms of on-line catalog use.

The relationship of journal and Internet use to JSTOR use

Examination of the frequency of JSTOR use among faculty aware of JSTOR (n=78) showed that 58% of the respondents had varying levels of use, while 42% reported no use. Using the frequency of JSTOR use as the dependent variable, the faculty who reported no use were censored on the dependent variable. The standard zero, lower bound Tobit model was designed for this circumstance (Tobin, 1958). Most important, by adjusting for censoring, the Tobit model allows inclusion of negative cases in the analysis of variation in frequency of use among positive cases, which greatly enhances degrees of freedom. Therefore, hierarchical Tobit regression analyses were used to examine the influence of demographic characteristics, journal use, search preferences, Internet use, and attitude toward computing on the frequency of JSTOR use. Independent variables used in these analyses were selected on the basis of significance in univariate Tobit regressions on the frequency of use variable. Table 9 summarizes the independent variables used in the multiple Tobit regression analyses.

Table 10 summarizes the results of the hierarchical Tobit regression of demographic, journal use, search preference, Internet use, and computing attitude variables on frequency of JSTOR use. The bottom line of Table 10 summarizes the log likelihood score for each model. Analysis of the change in log likelihood score between adjacent models gives a measure of the significance of independent variables added to the model. For example, in Model 1, the addition of the demographic variables failed to produce a significant change in the log likelihood score compared to the null model. By contrast, in Model 2, the addition of journal use variables produced a significant change in the log likelihood score compared to Model 1 -- suggesting that the addition of the journal use variables improved the fit in Model 2 over Model 1. Similarly, the addition of search variables in Model 3 and of Internet use variables in Model 4 both produced significant improvements in fit, but the addition of the computer attitude variable in Model 5 did not. Therefore, Model 4 was selected as the best model. From Model 4, the coefficients for gender, article copying, abstract reading, and searching on-line catalogs are all positive and significant. These results suggest that controlling for other factors, men were 0.77 points higher on frequency of JSTOR use than women, there was a 0.29 point increase in the frequency of JSTOR use for every point increase in the frequency of article copying, faculty who read article abstracts were 0.82 points higher on frequency of JSTOR use than faculty who didn't read abstracts, and there was a 1.13 point increase in the frequency of JSTOR use for every point increase in the frequency of on-line catalog searching. From Model 4, the coefficients for affiliation with an economics department and the number of paid journal subscriptions are both negative and significant. These results suggest that controlling for other factors, economists were 0.88 points lower on frequency of JSTOR use than historians, and there was a 0.18 point decrease in frequency of JSTOR use for every unit increase in the number of paid journal subscriptions.

Discussion

This study addressed five questions related to the impact of JSTOR: a) how faculty searched for information; b) which faculty used JSTOR; c) how journals were used d) how the Internet was used; and e) how journal use and Internet use correlated with JSTOR use.

Summary of findings

In terms of how faculty searched for information, results were consistent with earlier findings reported in the literature. Specifically, a strong majority of the faculty reported relying on citations from related publications, on colleagues, on electronic catalogs, and on browsing library shelves when seeking information. Faculty did not differ dramatically in selection of search strategies, except that Michigan economists were less likely to browse library shelves and less likely to search card catalogs.

In terms of JSTOR use, Michigan faculty were less likely to know about JSTOR than the five college faculty, and Michigan faculty were less likely to use JSTOR than the five college faculty. These results probably reflected the delayed rollout and availability of JSTOR at Michigan. Economists were more likely to use JSTOR than historians. Of the faculty who reported JSTOR use, frequency of use did not differ dramatically from frequency of use of a related, more traditional technology: on-line searching. That is, 58% of the faculty who used JSTOR said they used JSTOR once a month or more, while 69% of the faculty who did on-line searches reported doing searches once a month or more. Note however, that over twice as many faculty reported doing on-line searches (75%) as reported use of JSTOR (33%).

In terms of journal use, faculty did not vary greatly in their use of journal features, except that Michigan historians were less likely to use article abstracts, and Michigan economists were less likely to use book reviews. Economists and historians did not differ in the total number of journals used, however there was greater consensus among the economists about core journals. Specifically, two of the five economics titles included in JSTOR (the *American Economic Review* and the *Journal of Political Economy*) were cited among the top ten most frequently used journals by a majority of the economists, while four of the five titles (the two mentioned above plus the *Quarterly Journal of Economics* and the *Review of Economics and Statistics*) were cited among the top ten most frequently used journals by a majority of the Michigan economists. By contrast, only one of the five history titles included in JSTOR (the *American Historical Review*) was cited among the top ten most frequently used journals by a majority of the historians.

In terms of Internet use, the Michigan historians lagged their colleagues in economics at Michigan and the five college faculty. For example, the Michigan historians reported less use of email, the World Wide Web, ftp, and on-line databases than the other faculty. The economists were more likely to use ftp and more likely to use the World Wide Web than the historians. Faculty used on-line catalogs at similar rates.

In terms of factors correlated with JSTOR use, the tobit regressions showed that a model including demographic factors, journal use factors, search factors, and Internet use factors offered the best fit to the data on frequency of JSTOR use. The addition of the computer attitude variable did not improve the fit of this model. In the best fit model, gender, article copying, abstract reading, and searching on-line catalogs were all positively and significantly related to frequency of JSTOR use. Also from the best fit model, affiliation with an economics department and greater numbers of journal subscriptions were negatively and significantly related to frequency of JSTOR use.

Limitations of the study

These data represent a snapshot of faculty response to JSTOR at an extremely early stage in the evolution of the JSTOR system. In the spring of 1996, JSTOR had been available to the five college faculty for less than six months, while at Michigan, the system had not yet been officially announced to faculty. Therefore, the results probably underestimate eventual use of the mature JSTOR system. Further, as a survey study, self-reports of use were crude compared to measures that could have been derived from actual behavior. For example, it was intended to match use reports with automated usage statistics from the JSTOR Web servers, but the usage statistics proved too unreliable. Another problem was that the survey contained no items on the frequency of traditional journal use. Therefore, it is unknown whether the low use of JSTOR reported by the faculty reflected dissatisfaction with the technology or simply a low baserate for journal use. Finally, the faculty at Michigan and at the five colleges were atypical in the extent of their access to the Internet and in the modernity of their computing equipment. Faculty with older computers and slower network links would probably be even less likely to use JSTOR.

Implications for the JSTOR experiment

Although extremely preliminary, these early data suggest trends that merit further exploration as JSTOR expands. First, it is encouraging to discover that among faculty who have used JSTOR, rates of use are already comparable to rates for use of on-line searching -- a technology that pre-dates JSTOR by two decades. It will be interesting to see if JSTOR use grows beyond this modest level to equal the use of key Internet applications, like email and Web browsing. Second, there appear to be clear differences in journal use across disciplinary lines. For example, economists focus attention on a smaller set of journals than is the case in history. Therefore, it may be easier to satisfy demand for on-line access to back archives in fields that have one or two flagship journals than in more diverse fields where scholarly attention is divided among dozens of journals. This may lead commercial providers of back archive content to ignore more diverse disciplines at the expense of easier to service focused disciplines. Finally, the negative correlation between the number of journal subscriptions and JSTOR use suggests the possibility of a substitution effect (i.e., JSTOR for paper). However, the significance of this correlation is difficult to determine, since there is no way to know the direction of causality in a cross-sectional study.

References

Broadbent, E.A. (1986). Study of humanities faculty library information seeking behavior. *Cataloging and Classification Quarterly*, *6*, 23-37.

Carley, K., & Wendt, K. (1991). Electronic mail and scientific communication: A study of the SOAR extended research group. Knowledge: Creation, Diffusion, Utilization, 12, 406-440.

Finholt, T.A., & Olson, G.M. (1997). From laboratories to collaboratories: A new organizational form for scientific collaboration. *Psychological Science*, *8*, 28-36.

Fox, E.A., Akscyn, R.M., Furuta, R.K., & Leggett, J.J. (Eds.). (1995). Digital libraries [Special issue]. Communications of the ACM, 38(4).

Garvey, W.D. (1979). *Communication: The essence of science*. Toronto: Pergamon Press.

Garvey, W.D., Lin, N., & Nelson, C.E. (1970). Communication in the physical social sciences. *Science*, *170*, 1166-1173

Kling, R., & Covi, L. (1996). Electronic journals and legitimate media. *The Information Society*, *11*, 261-271.

Lougee, W.P., Sandler, M.S., & Parker, L.L (1990). The Humanistic Scholars Project: A study of attitudes and behavior concerning collection storage and technology. *College and Research Libraries*, *51*, 231-240.

Odlyzko, A. (1995). Tragic loss or good riddance? The impending demise of traditional scholarly journals. *International Journal of Human-Computer Studies*, *42*, 71-122.

Olsen, J. (1994). *Electronic journal literature: Implications for scholars*. Westport, CT: Mecklermedia.

Sabine, G.A., & Sabine, P.L. (1986). How people use books and journals. *Library Quarterly*, *56*, 399-408.

Schuegraf, E.J., & van Bommel, M.F. (1994). An analysis of personal journal subscriptions of university faculty. Part II: Arts and professional programs. *Journal of the American Society of Information Science*, *45*, 477-482.

Simpson, A. (1988). Academic journal usage. *British Journal of Academic Librarianship*, *3*, 25-36.

Stenstrom, P., & McBride, R.B. (1979). Serial use by social science faculty: A survey. *College and Research Libraries*, *40*, 426-431.

Taubes, G. (1993). Publication by electronic mail takes physics by storm. Science, 259, 1246-1248.

Tobin, J. (1958). Estimation of relationship for limited dependent variables. Econometrica, 26, 24-36.

Walsh, J.P., & Bayma, T. (1997). Computer networks and scientific work. In S.B. Kiesler (Ed.), *Culture of the Internet*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Table 1
Percentage of faculty by search strategy, type of institution and discipline (n=151[a])

Search strategies	University of Michigan Economics (n=44)	University of Michigan History (n=54)	Five colleges Economics (n=25)	Five colleges History (n=28)
Use citations from related publications	84%	96%	100%	100%
Consult a colleague	93%	85%	96%	89%
Search electronic catalogs for a known item	80%	89%	88%	89%
Browse library shelves	46%a	83%	72%	86%b
Browse electronic catalogs	57%	56%	80%	79%
Use electronic indexes	59%	59%	84%	64%
Use printed indexes	34%	57%	64%	82%
Search card catalogs for a known item	14%a	32%	17%	39%b
Browse departmental collections	48%a	11%	20%	4%b
Browse card catalogs	2%	20%	24%	25%

Note: Means with different subscripts differ significantly at $p < .01$ in the Tukey honestly significant difference test. [a] 9 cases were unusable due to incomplete data.

Table 2
Percentage of faculty by frequency of JSTOR use, type of institution and discipline (n=147[a])

Frequency of use	University of Michigan Overall (n=93)	University of Michigan Economics (n=43)	University of Michigan History (n=50)	Five colleges Overall (n=54)	Five colleges Economics (n=26)	Five colleges History (n=28)
never[b]	73%	60%	84%	58%	54%	61%
once a year	12%	17%	8%	17%	15%	18%
once a month	9%	14%	4%	14%	19%	10%
once a week	6%	9%	4%	11%	12%	11%
daily	0%	0%	0%	0%	0%	0%

Note: [a] 13 cases were unusable due to incomplete data. [b] The "never" category also includes faculty who were unaware of JSTOR.

Table 3
Percentage of faculty by frequency of on-line searching, type of institution and discipline (n=147[a])

Frequency of searches	University of Michigan Overall (n=93)	University of Michigan Economics (n=43)	University of Michigan History (n=50)	Five colleges Overall (n=54)	Five colleges Economics (n=26)	Five colleges History (n=28)
never	24%	24%	24%	15%	11%	18%
once a year	25%	28%	22%	24%	16%	32%
once a month	25%	22%	28%	26%	34%	18%
once a week	23%	19%	26%	30%	35%	25%
daily	3%	7%	0%	6%	4%	7%

Note: [a] 13 cases were unusable due to incomplete data.

Table 4
Percentage of faculty by use of journal features, institution and discipline (n=159[a])

Journal feature	University of Michigan Economics (n=47)	University of Michigan History (n=58)	Five colleges Economics (n=26)	Five colleges History (n=28)
Articles	96%	98%	100%	100%
Table of Contents	81%	86%	100%	96%
Bibliographies	60%	71%	89%	82%
Book Reviews	49%b	98%a	85%a	100%a
Article Abstracts	81%a	31%b	89%a	61%a
Editorials	13%	24%	35%	43%

Note: Means with different subscripts differ significantly at $p < .01$ in the Tukey honestly significant difference test.

[a] 1 case was unusable due to incomplete data.

Table 5
Number of journals used, number of paid subscriptions, and number of JSTOR target journals by institution and discipline (n=160)

Journals used	University of Michigan Economics (n=48)	University of Michigan History (n=58)	Five colleges Economics (n=26)	Five colleges History (n=28)
Total	8.6	8.9	8.4	8.7
Number that are paid subscriptions	3.7	4.6	4.0	3.6
Number that are JSTOR target journals	3.1a	1.6b	2.5	1.9b

Note: Means with different subscripts differ significantly at $p < .01$ in the Tukey honestly significant difference test.

Table 6
Percentage of economics faculty ranking JSTOR economics journals as top five most frequently used, next five most frequently used, and not used (n=74)

	University of Michigan (n=48)			Five colleges (n=2		
Journal	Top five	Next five	Not used	Top five	Next five	Not
American Economic Review	79%	6%	15%	66%	15%	19%
Journal of Political Economy	52%	10%	38%	32%	26%	42%
Quarterly Journal of Economics	41%	15%	44%	16%	26%	58%
Econometrica	26%	30%	44%	8%	15%	77%
Review of Economics and Statistics	18%	28%	54%	12%	34%	54%

Table 7
Percentage of history faculty ranking JSTOR history journals as top five most

frequently used, next five most frequently used, and not used (n=86)

Journal	University of Michigan (n=58)			Five colleges (n=2		
	Top five	Next five	Not used	Top five	Next five	Not
American Historical Review	44%	19%	37%	58%	24%	18%
Journal of American History	31%	6%	63%	39%	4%	57%
Journal of Modern History	15%	10%	75%	18%	11%	71%
William and Mary Quarterly	13%	6%	81%	15%	3%	82%
Speculum	9%	3%	88%	11%	10%	79%

Table 8
Mean frequency of computer application use over direct connection (high speed network) by institution and by discipline (n=158[a])

Computer application	University of Michigan Economics (n=47)	University of Michigan History (n=57)	Five colleges Economics (n=26)	Five colleges History (n=28)
Email	4.9a	3.3b	5.0a	4.7a
On-line Catalogs	3.3	2.8	3.6	3.7
On-line Databases	2.3	1.6a	2.9b	2.1
World Wide Web	3.9a	1.8b	4.2a	2.9c
File Transfer Protocol (ftp)	3.4a	1.4b	2.7a	1.7b

Note: Frequency of use was reported on a 5-point scale (1 = *never*; 2 = *2 or 3 times per year*; 3 = *monthly*; 4 = *weekly*; 5 = *daily*). Means with different subscripts differ significantly at $p < .01$ in the Tukey honestly significant difference test.

[a] 2 cases were unusable due to incomplete data.

Table 9
Descriptive statistics for faculty aware of JSTOR (n=78)

Variable	Mean	STD
at Michigan	49%	--
in economics	54%	--
male	82%	--
years since degree	17.2	11.5
copies articles	3.09	0.91
puts articles on reserve	2.73	1.15
reads abstracts	68%	--
total # subs., JSTOR	2.5	1.5
total # subs., all	8.8	1.96
# paid subs.	4.04	2.43
use on-line indexes	60%	--
search on-line catalog	85%	--
browse on-line catalog	65%	--
frequency of on-line catalog use	3.47	1.25
frequency of on-line database use	2.33	1.31
frequency of WWW use	3.47	1.62
frequency of ftp use	2.39	1.42
attitude toward computing	3.52	0.70
frequency of JSTOR use	2.05	2.09

Table 10
Tobit regression on frequency of JSTOR use among faculty aware of JSTOR (n=78)

Variable	Model 1	Model 2	Model 3	Model 4
Constant	0.56	-2.45*	-3.89***	-3.86***
at Michigan	-0.11	.28	.47	.47
in economics	0.20	-.73	-.48	-.88*
male	.77	.82*	.91**	.77*
years since degree	-0.04**	-0.02	-0.00	0.00
copies articles		.29	.28	.29*
puts articles on reserve		.28*	.33**	.24
reads abstracts		1.38***	1.22***	.82**
total # subs., JSTOR		.27*	.26*	.21
total # subs., all		0.03	-0.02	-0.02
# paid subs.		-.17**	-.16**	-.18**
use on-line indexes			.37	.22
search on-line catalog			1.34**	1.13*
browse on-line catalog			-0.02	-.15
frequency of on-line catalog use				0.02
frequency of on-line database use				0.02
frequency of WWW use				.22
frequency of ftp use				.20
attitude toward computing				
-log likelihood	111.94	98.08	93.56	89.31
Chi-square	6.72	27.72***	9.04**	8.5*

Note: -log likelihood for the null model = 115.30 $^{*} = p < .10$; $^{**} = p < .05$; $^{***} = p < .01$

Figure 1. Cumulative percentage of on-line searchers versus JSTOR users, by frequency of use and type of institution (n=147)

FOOTNOTES:

[1] At the time of this study, the Department of Economics at the University of Michigan maintained an extensive departmental library with support from the central library. This departmental collection is no longer supported.

[2] This combines the 44% of the faculty who were unaware of JSTOR with the 23% of the faculty who were aware of JSTOR, but did not use it.

……

For additional information about the conference, or The Andrew W. Mellon Foundation's scholarly communication initiatives, please contact Richard Ekman. For additional information about ARL or this web site contact Patricia Brennan, ARL Program Officer at (202) 296-2296.

Return to Office of Scholarly Communication Home Page

ARL Home ARL Scholarly Communication and Technology Home Page

© Association of Research Libraries, Washington, DC

Web Design by Angelo F. Cruz
Maintained by ARL Web Administrator
Last Modified: August 29, 1997

Scholarly Communication and Technology

Conference Organized by The Andrew W. Mellon Foundation

at Emory University
April 24-25, 1997

Copyright © of the papers on this site are held by the individual authors or The Andrew W. Mellon Foundation. Permission is granted to reproduce and distribute copies of these works for nonprofit educational or library purposes, provided that the author, source, and copyright notice are included on each copy. For commercial use, please contact **Richard Ekman** at the The Andrew W. Mellon Foundation.

Session #4 Patterns of Usage

Patterns of Use for the Bryn Mawr Reviews

Richard Hamilton and Paul Shory
Professor of Greek
Bryn Mawr College

. . . .

Historical Background

Bryn Mawr Reviews (BMR) produces two electronic review journals *Bryn Mawr Classical Review* (BMCR), which also comes out in paper and was started at the end of 1990 and *Bryn Mawr Medieval Review* (BMMR), started in 1993. After about two years of activity BMMR became dormant and toward the end of 1996 both location and management were shifted [1]; since then it has become tremendously active, at one point even surpassing BMCR in its monthly output.[2] The comparisons below should be considered with this in mind.

Data

We have two sets of users: subscribers and gopher hitters.[3] As data from the former we have subscription lists, which are constantly updated, and periodic surveys that we have conducted; for the latter we have monthly reports of gopher hits and gopher hitters (but not what the hitters hit). In considering this data our two main questions have been: how are we doing; how can we afford to keep doing it.

A. Gopher Reports

Our analysis of the monthly gopher reports has concentrated on the hitters rather than the hits. After experimenting rather fruitlessly in 1995 with microanalysis of the data from the Netherlands and Germany hitter by hitter month by month for a year, we decided to collect only the following monthly figures:

total # users

total by address (country, edu, com etc.)

list of top hits (those reviews that received 15+ hits/month and are over a year old[4])

list of top hitters (those who use the system 30+/month).

Analysis of the total users shows that use has levelled off at a peak of about 3800 users a month (see appendix). With a second full year of gopher use to study we can see the seasonal fluctuation more easily. The one area of growth seems to be non-English foreign sites. If we compare the top hitters in the first ten months of 1995 with the comparable period in 1996 we find that the total increased only 5% but the total number of non-English heavy users increased 120%. Three countries were among the heavy users in both 1995 and 1996 (France, Germany, Netherlands); two appeared only in 1995 (South Africa, Taiwan) and eight only in 1996 (Brazil, Italy, Ireland, Poland, Portugal, Russia, Spain, Venezuela).

Chart 1: BMCR/BMMR Top Hitters (30+ hits a month)

	US	English	Non-English	Total
1995	47	8	5	60
1996	42	10	11	63

In terms of number of total users from 1995 to 1996 there was an overall increase of 10.8%, though the increase among US users was only 9.1%. Conversely, most foreign countries if anything showed a marked increase in total use over the ten months of 1996 vs 1995 (see appendix): Argentina 16 to 27, Australia 542 to 684, Brazil 64 to 165, Denmark 80 to 102, Spain 107 to 197, Greece 41to 80, Ireland 50 to 69, Israel 89 to 108, Italy 257 to 359, Japan 167 to 241, Korea 26 to 40, Netherlands 273 to 315, Portugal 16 to 26, Russia 9 to 27, (former) USSR 13 to 20, and South Africa 63 to 88. On the other hand, Iceland went from 22 to 8, Malaysia from 30 to 21, Mexico from 68 to 56, Sweden from 307 to 250, and Taiwan from 24 to 14. Also, among US users there was a large drop in edu from 7073 to 5962 and a corresponding rise in net from 1570 to 4118, perhaps because faculty members are now using commerical providers for home access.[5]

In the analysis of top hits a curious pattern emerges: BMMR starts out with many more top hits despite there being a much smaller number of reviews (about 15% of BMCR's number) but toward the end of 1995 the pattern shifts. BMMR dominates at the beginning but drops when

BMMR becomes inactive.

Chart 3: Favorite Reviews (reviews at least one year old that received 15+ hits/month)

month	BMMR	BMCR
1/95	2	1
2/95	15	11
3/95	10	6
4/95	2	3
5/95	5	5
6/95	16	20
7/95	3	1
8/95	12	14
9/95	41	116
10/95	46	170
1/96	38	81
2/96	14	69
3/96	15	74
4/96	19	50
5/96	6	25
6/96	9	13
7/96	7	16
8/96	8	19
9/96	20	48
10/96	14	54

The shift is easily explained since it occurs about the time BMMR was becoming inactive, but the original high density is still surprising.[6] Likewise medieval books receive noticeably more attention: 32 medieval titles made the top hits list 116 times (avg 3.6) while 81 classical titles made the list only 219 times (avg 2.7), despite including two blockbuster titles, Amy Richlin's *Pornography and Representation* (10x) and John Riddle's *Contraception and Abortion* (14x).[7]. My guess is that medievalists, being more widely dispersed in interests and location, have found the Net more important than have classicists, who are mostly located in a classics department and whose professional work is more circumscribed (and has a longer history).

B. Subscriptions

Subscriptions to the e-journals continue to grow at a rate of 5% per quarter, though there is considerable seasonal fluctuation:

Chart 4: Subscriptions

	3/95	6/95	9/95	3/96	6/96	10/96
BMCR	1072	1067 (-.4%)	1135 (+ 6%)	1253 (+10%)	1273 (+2%)	1317 (+ 3%)
BMMR	711	755 (+ 6%)	865 (+13%)	931 (+ 8%)	964 (+4%)	995 (+ 3%)
joint	568	562 (- 1%)	599 (+ 7%)	672 (+12%)	685 (+2%)	770 (+12%)
total	2351	2384 (+ 1%)	2599 (+ 9%)	2856 (+10%)	2922 (+2%)	3082 (+ 5%)

Looking more broadly we see a steady slowdown in growth of all but the joint subscriptions:

	9/93	9/94	9/95	10/96
BMCR	651	882 (+35%)	1135 (+29%)	1317 (+16%)
BMMR	257	498 (+94%)	865 (+74%)	995 (+15%)
joint	261	460 (+76%)	599 (+30%)	770 (+29%)

If we look at the individual locations, we find again that while the US subscriptions continue to grow, they are becoming steadily less of the whole, going from 77% of the total in 1993 to 68% in 1996. English-speaking foreign countries have remained about the same percentage of the whole; it is non-English speaking foreign contries that have shown the greatest increase, going from 4% of the total in 1993 to 13% of the total in 1996.

Chart 5: BMCR Subscribers

	1993	1994	1995	1996
total	730	1019	1130	1349
edu	529	701	703	779
com	22	44	72	103
gov	3	6	4	4
mil	2	2	2	2
org	5	6	7	12
net	3	5	8	17
US total	564 (77%)	764 (75%)	796 (70%)	917 (68%)
foreign total	154	254	332	428
ca	58	87	106	114
uk	31	45	57	77
au	21	33	38	43
nz	4	6	7	6
za	8	12	14	18
ca/uk/au/nz/za	122 (17%)	183 (18%)	222 (20%)	258 (19%)
non-English	32 (4%)	71 (7%)	110 (10%)	170 (13%)

de	5	11	16	27
nl	7	10	16	24
ie	1	4	5	5
fi	3	8	9	12
br	0	2	2	2
fr	1	4	7	9
es	0	0	1	3
it	2	4	7	17
hu	0	2	2	2
ve	1	1	1	1
se	3	4	6	7
gr	0	1	3	8
il	2	6	11	14
dk	1	1	1	0
no	3	4	4	4
kr	0	0	1	1
be	0	2	5	7
us	0	2	2	4
jp	1	2	3	4
ch	1	2	4	12
pt	0	0	1	1
at	0	0	1	2
hk	0	1	1	1
my	0	0	1	1
tr	0	0	1	1
pl	0	0	0	2

C. Subscriber Surveys

As opposed to the gopher stats, which give breadth but little depth, our surveys offer the opportunity for deeper study of our users but at the expense of breadth. We cannot survey our subscribers too often or they will not respond.[8] A further limitation is that we felt we could not survey those who take both BMCR and BMMR, a significant number, without skewing the results since many subscribers lean heavily toward one journal or the other and the journals are significantly different in some ways. So far we have conducted five surveys:

1) a 20 question survey to BMCR subscribers November 1995

2) a 21 question survey to BMMR subscribers in February 1996

3) a 2 question survey to all subscribers in October 1996[9]

4) a 15 question survey to all BMCR reviewers whose e-mail addresses we knew in January 1997.

5) a 2 question survey to those who have cancelled subscriptions in the past year (March 1997).

Here is the subscriber profile as revealed in the surveys:

	BMCR	BMMR
male	72.3%	50.1%
female	25.3	44.8
AB	5.5	9.6
ABD	12.8	18.0
PhD	66.6	49.3

faculty	65.0	44.2
adjunct, research	7.0	6.5
grad student.	15.1	23.7
undergrad	.8	2.3
check e-mail daily	90.3	85.9
read review on screen	66.8	63.9
print immediately	6.5	5.9
read on screen to decide	24.5	27.3
never/rarely delete w/o reading	83.1	85.4
made printed copy sometimes/often	56.9	51.9
copies on disk sometimes/often	51.7	50.7
have used gopher	42.0	15.8
reviewed for journal	25.1	9.6
heard reference to journal	70.0	31.0
finish a few reviews	42.0	19.7
finish many/most reviews	53.5	64.8
finish almost all	3.1	13.2
review useful for teaching	53.8	41.1
review useful for research	87.2	78.9
willing to pay $5 subscription	66.8	50.1

Many of the differences are easily explained by the chequered history of BMMR or by the differing natures of the two readerships.[10] I doubt many will be surprised that medievalists are more often female and less often faculty. The paucity of reader-reviewers of BMMR reflects the paucity of BMMR reviews. To me the most surprising statistic is how few of subscribers to either journal have used gopher.

The key question of course is willingness to pay for subscription, and with that in mind we did some correlation studies for the BMCR survey, first seeing with what variables there was a correlation with a willingness to pay $5 for a subscription. We found positive correlation with

ever found review useful for teaching (.0004 likelihood of a chance correlation)

ever found review useful for research (.00006)

ever hear a reference to BMCR (.00001)

ever written a review for BMCR (.00089)

Some further correlations were found:

start to read many or most reviews// heard a reference to BMCR (.00014)

willing to review// heard a reference to BMCR (.00003)

get paper BMCR// have written review (.00003)

have written review// will write in future (.0000)

will write in future// library gets BMCR (.00007)

PhD// willing to review (.00001).

A follow-up two question survey done in October 1996 asked whether subscribers would prefer to pay for e-mail subscription, or to receive advertisements from publishers or to cancel. 14% preferred to pay, 82% to receive advertisements and 4% to cancel.

Our most recent survey, of those who had for one reason or another dropped from the list of subscribers, revealed that almost a third were no longer valid addresses and so were not true cancellations. Of those who responded almost half (40, 44%) of the unsubscriptions were only temporary. The reason for cancellation was rarely the quality of the review.

Chart 7: BMCR Unsubscriber Survey (those who unsubscribed 1/96-2/97)

317 total: 103 address no longer valid; 91 responses

identity

15 unaffiliated with academic institution

46 faculty (4 retired, 9 adjunct or research)

7 librarians

8 students (2 undergraduates)

7 other

reasons (faculty # in parenthesis)

2 never subscribed (1)

2 never meant to unsubscribe (1)

16 unsubscribed from old, subscribed to new address (14)

15 suspended subscription while away (9+1)

22 decided reviews not sufficiently relevant to interests (6+2)

2 decided review quality not high enough (+1)

11+3 too much e-mail (6+3)

7+1 no longer have time to read reviews (+2)

7+1 other (5 shifted to BMR, 1 to BMCR, mistake) (4+1)

question	unaffiliated	faculty	librarian	student	ot
not relevant	8	6+2	1	2	2
too much mail	2	7	-	2	-
no time	4	+2	-	-	1
total	14	13+4	1	4	3

Conclusions

If we return to our two questions: progress and cost recovery, we can see that our progress is satisfactory but cost-recovery is still uncertain.

BMCR is growing at the rate of 30% a year.[11] The major American Classics organization (The American Philological Association) has a membership of about 3,000 members and so one may estimate the total world population of Classicists as somewhere between 7,000 and 10,000. If half of them have access to computers, BMCR presently reaches somewhere between 22% and 32% of its total market. At its present rate of growth, it will saturate its market in five years. It is much more difficult to estimate the total world market for BMMR, but it is certainly greater than that for BMCR, so with its present growth rate of perhaps 30%[12] it will take somewhat longer to reach saturation.

BMCR costs are about $4,000/year for over 700 pages of reviews. About half the cost goes for producing the paper version and we anticipate costs of between $1,500 and $2,000 per year for preparing reviews for the Web.[13] Uncompensated editorial time averages 34 hours/month. So, total out-of-pocket expenses could be as high as $6,000 if the paper version continues and if mark-up continues to be done by hand. A third possible reduction in costs besides elimination of the paper version and automatic mark-up is a "fast-track" system whereby the review never leaves the net: it is e-mailed to the editor who sends it to a member of the editorial board and when the two have made changes it is sent back to the reviewer for approval and then published on the net. The great advantage for the reviewer is that this cuts publication time by a month; the disadvantage is that the reviewer is asked to do some simple mark-up on the text before sending it.

Possible revenue sources include: advertising, subscriptions and institutional support. As we have seen, our subscribers much prefer receiving advertising to paying for a subscription, but we have no idea how successful we will be in attracting advertising.[14] At the Conference, Hal Varian suggested we try to arrange something with Amazon Books, and we will. We will not consider charging for subscriptions until BMCR is on the Web; at that point we could charge for timely delivery of the review, perhaps several months before universal access. We also want to wait for wide acceptance of a simple electronic cash transfer system. Institutional support seems to us the most obvious way to cover costs since the College gets considerable exposure for what seems to us a small cost.

BMCR/BMMR total gopher users

	96/1	96/2	96/3	96/4	96/5	96/6	96/7	96/8
Total	3622	3809	3778	3510	2887	2600	2687	2893
#	1015	1216	1235	1123	944	819	843	873
ae	1	1		1		1		
ai		1						
am			1					
ar		5	2	3	1	2	1	2
arpa	4	3	1	6	2	2	2	1
at	14	9	13	10	10	10	8	11
au	67	52	72	71	72	61	51	82
aw					1			
bb								
be	12	15	11	8	6	7	14	10
bg	1						1	
bh							1	1
bm			1	1				
bn			1					1
br	13	10	17	10	17	18	17	25
ca	165	154	153	144	101	90	82	100
ch	13	7	5	8	10	7	11	12
cl	3	2	2	1	2	4	1	3
cn				1	1	1	1	2
co			1	1	2			
com	503	466	453	362	322	316	329	356
cr	2		1	1	1	1	2	
cs								
cy		1						
cz	3	4	3	2	1	1	2	1
de	84	65	71	60	58	61	56	60
dk	11	17	8	7	7	4	8	13
do		1	1	1				1
dz								
ec								
edu	684	742	650	716	503	430	460	461
ee	3	2	2	1	4	2	1	
eg				2				
es	13	13	22	22	22	15	19	23
fi	27	23	23	18	13	9	23	14
fj	36	1						
fr		31	34	32	26	27	28	23
gb		1	1				1	
ge	1							
gi								
gov	52	60	52	49	43	29	29	29
gr	11	12	6	5	7	8	7	7
gt			1					
gu								
hk	1	4	2	2	1	1		2
hr	2	1	2	1	3	3		2
hu	6	1	3	1	2		5	6
id		2		1	2			
ie	7	9	6	8	9	4	4	3
il	11	11	18	15	7	7	8	10
in				1				1
int						1		
is	6	4	2	2	1	1		1
it	37	32	34	35	33	23	34	41
jm						1		
jo								

jp	31	31	23	24	16	25	20	16
kr	2	3	5	7	4	3	4	1
kw	1							
lb								
lt	2	1		1		1	1	
lu	2	3	1				1	
lv	1	2	1					
ma	1	1						
md						1		
mil	24	27	20	20	18	14	20	9
mo	1							
mt		1		2	2			
mx	7	6	10	3	6	7	3	4
my	2	3	2	5	2	1	2	1
na						1		
net	381	439	480	397	319	316	326	406
nl	43	38	23	37	28	24	33	30
no	24	18	22	16	15	9	10	12
nz	10	9	14	16	10	9	10	18
org	40	34	42	42	43	37	33	28
pa								
pe	1		1	1		1	1	1
ph			2	1		1		2
pk			1					
pl	5	4	3		3	7	3	1
pt	3	2	2	2	3	1	2	3
ro	1	1	1	1	1	1		1
ru	1	1	2	3	1	3	3	4
se	40	31	20	23	21	17	21	25
sg	5	4	2	4	4	4	2	2
si		3	3	1	1	2	3	2
sk				2				1
su	3	1	2	3	2	2	2	1
sz								
th	1	1	1		1	1	3	2
tr	7	2	4	2	1	1	1	3
tw	2	1	3		1	3	2	1
ua	1			1				1
uk	100	83	90	82	82	92	82	88
us	60	72	81	68	61	42	39	43
uy	2	1		1		1	1	1
ve						1	1	1
ws								
wst	1							
za	9	8	7	13	8	6	9	8

FOOTNOTES:

[1] It has as of May 7 become *The Medieval Review* (TMR).

[2] The output by month (4/95-3/97) is:

```
bmmr  10  17   5   8   4   3   5  11   6   7   4   6   1   4   1   6   6   9   8
bmcr  15  14  19  13  11  29  26  17  27  12  14  15  37   7  12  19  17  28  14
```

[3] Since May 7, BMMR (=TMR) has been on the Web, and this will eventually provide valuable data to compare with the BMCR gopher data.

[4] Naturally, new reviews are visited often; we are trying to isolate those of enduring value.

[5] Likewise, mil(itary) dropped from 310 to 186; gov(ernment) from 819 to 409.

[6] The explosive growth in 9/95 and 10/95 was only temporary.

[7] The difference would be even more pronounced had I not excluded books that appeared on the list only once. In 1996 the gap virtually disappears: 31 medieval titles (total number of titles 53) made the list 126 times (avg. 4.1) while 93 classical titles (total number of titles 169) made the list 360 times (avg. 3.9).

[8] As is our response rate is only in the 30-40% range.

[9] Unfortunately the survey was worded as if only for BMCR subscribers, but even so the response rate was about 35%.

[10] We found similar differences in a pilot comparison of qualitative differences in the two journals done by two advanced graduate students (one a Classicist one a Medievalist) in the summer of 1995. They concluded that the major differences stem from the scholarly orientation of either discipline not from their media (i.e., Classicists criticize at a microscopic level, assume in-depth acquaintance with a given text). The reviews are longer and the number of typographical errors is much greater in BMMR but other differences seemed to be personal (tone of the review, footnotes and additional bibliography, organization, amount of direct quotation).

[11] Combined BMCR and joint figures= 912 for 1993, 1342 for 1994 (+47%), 1734 for 1995 (+29%) and 2264 for 1996 (+30%).

[12] Combined BMMR and joint figures= 518 for 1993, 958 for 1994 (+85%), 1464 for 1995 (+53%) and 1765 for 1996 (+21%). We have already seen an increase since BMMR relocated (3/97 = 1985, c.30% annually) and we may expect a considerable bump after official unveiling of TMR at the annual conference in May (and the introduction of the website).

[13] BMMR has found it takes 35 minutes on average to SGML a review.

[14] So far only Princeton University Press (of the eight contacted) has signed up for e-advertising.

……

For additional information about the conference, or The Andrew W. Mellon Foundation's scholarly communication initiatives, please contact Richard Ekman. For additional information about ARL or this web site contact Patricia Brennan, ARL Program Officer at (202) 296-2296.

Return to Office of Scholarly Communication Home Page

ARL Home ARL Scholarly Communication and Technology Home Page

© Association of Research Libraries, Washington, DC

Web Design by Angelo F. Cruz
Maintained by ARL Web Administrator
Last Modified: August 29, 1997

Scholarly Communication and Technology

Conference Organized by The Andrew W. Mellon Foundation

at Emory University
April 24-25, 1997

Copyright © of the papers on this site are held by the individual authors or The Andrew W. Mellon Foundation. Permission is granted to reproduce and distribute copies of these works for nonprofit educational or library purposes, provided that the author, source, and copyright notice are included on each copy. For commercial use, please contact **Richard Ekman** at the The Andrew W. Mellon Foundation.

The original version of this web document can be seen at <http://community.bellcore.com:80/lesk/schcom/schcom.html>

Session #4 Patterns of Usage

Digital Libraries: A Unifying or Distributing Force?

Michael Lesk
Division Manager
Computer Science Research
Bellcore

Digital Libraries: A Unifying or Distributing Force?

Michael Lesk

For presentation at *Scholarly Communication and Technology*, a conference sponsored by the Andrew W. Mellon Foundation, Atlanta, Georgia (April 24, 1997).

Abstract

What kinds of communities will digital library technology produce? The Web seems much more popular than electronic journals. Does this mean that surfing will replace literature reading, and that nerds building HTML hierarchies will supplant publishers? Will this means that the universities lose control of the quality of what their students read? Will the ability to do more research in one's dorm room mean that students do not talk to one another at all, that they talk to people somewhere else in the world, or that they talk to their roommates more than ever, perhaps about how to use the computer system?

Digital information threatens our ideas of locality: will the association of students with a particular university, let alone university library, survive the Web? Might we find that online references and online multimedia lectures would produce the `virtual university of the United States' and if so would we want that? Universities serve a variety of social functions which the Web can augment or diminish, depending on our actions. The Web also may threaten our ideas of quality in scholarship. This paper addresses potential consequences of the change to digital information, and suggests that universities can cope by being more proactive in their use of the Web for reward and communication.

Introduction

There are several future trends that everyone seems to agree upon. They include

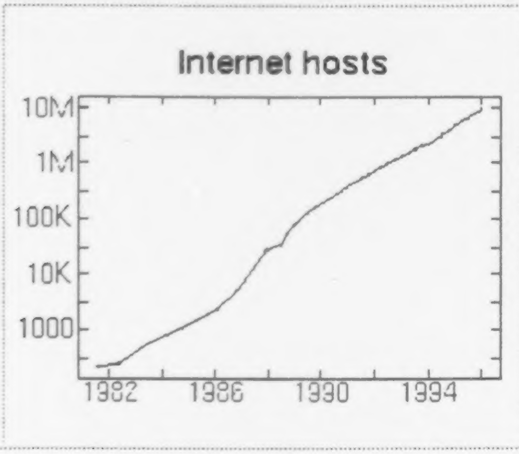
> widespread availability of computers for all college and university students and faculty;
> general substitution of electronic for paper information;
> library purchase of access to scholarly publications, rater than physical copies of them.

Early steps in these directions have been followed by many libraries. Much of this has taken the form of digitization. Unfortunately some of the digitized material is not used as much as we would like. This may reflect the choice of the material to convert; realistically 19th century books which have never been reprinted or microfilmed may have been obscure for good reasons and will not be used much in the future. But some more general problems with the style of much electronic library material suggest that the difficulties may be more pervasive.

The Web

The primary means today whereby people gain access to electronic material is over the World Wide Web. The growth of the Web is amply documented at http://www.cyberatlas.com and similar sites. Predictions for the number of Web users world wide in the year 2000 run up to 1 billion [Negroponte 1995]; students have the highest Web usage of any demographic group, with about 40% of them in 1996 showing medium or high Web usage; and people have been predicting the end of paper libraries since at least 1964 [Samuel 1964]. Web surfing appears to be substituting for TV viewing and CD-ROM purchasing, taking its share of approximately 7 hours per day that the average American spends dealing with media of all forms. Advertisers are lining up to investigate Web users and find the best way to send product messages to them [Hoffman 1996].

The table below shows the growth of Web hosts just in the last three years (from Cyberatlas and Network Wizards):



Internet hosts

Date	Number
Jan-96	9,472,224
Jul-95	6,642,000
Jan-95	4,852,000
Jul-94	3,212,000
Jan-94	2,217,000
Jul-93	1,776,000
Jan-93	1,313,000
Jul-92	992,000

Online Journals and the Web

Following the move of information to digital form, there are many experiments with online journals. Among the best known projects of this sort are the TULIP project of Elsevier [Hunter 1996] and the CORE project of Cornell, the American Chemical Society, Bellcore, Chemical Abstracts, and OCLC. These projects achieved more or less usage, but none of them approached the degree of epidemic success shown by the Web. The CORE project, for example, logged 87,000 sessions of 75 users, but when we ended access to primary chemical journals at Cornell, nobody stormed the library demanding the restoration of service. You can imagine what would happen if the Cornell administration were to cut access to the Web.

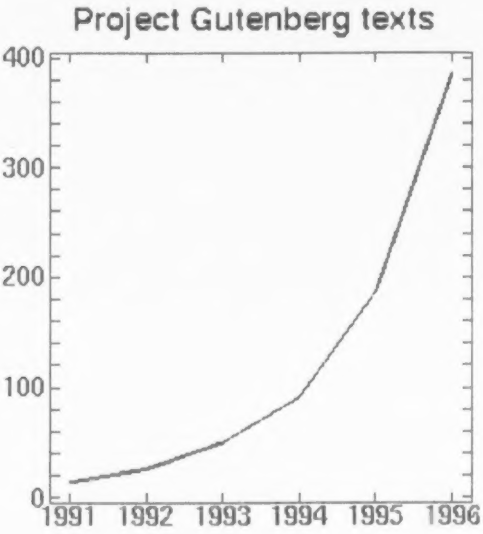
In the CORE project (see Entlich 1997), the majority of the usage was from the Chemistry and Materials Science departments. They provided 70% of active users and 86% of all sessions with the journals. There are various other departments at Cornell which use chemical information (Food Sciences, Chemical Engineering, etc.) but make less use of the online journals. Apparently the overhead of starting to use the system and learning its use discouraged those for whom it was not their primary interest. Many of the users printed out articles rather than read them online; about one article was printed for every four viewed, and people tended to print an article rather than flip through the bitmap images. People accessed articles through both browsing and searching, but they read the same kinds of articles they would have read otherwise, rather than changing their reading habits.

Some years ago the CORE project had compared the ability of people to read bitmaps compared with reformatted text, and found that people could read screen bitmaps just as fast as new text [Egan 1991]. Yet, in the actual use of the journals, the readers did not seem to like the page images. The Scepter interface provided a choice of page image or text format, and readers only looked at about one image page in every four articles. This suggests that despite assertions by some chemists in early interviews that they particularly liked the layout of ACS journal pages, for viewing online they prefer reformatted text to images of those pages, even though they can read either at the same speed. The Web-like style is preferred for online viewing.

Perhaps it is not surprising that the Web is more popular than scientific journals. After all, *Analytical Chemistry* has never had the circulation among undergraduates of *Time* or *Playboy*. But the Web is not being used only to find out sports scores or other non-scholarly activities (30% of all *Alta Vista* queries are about sex) [Weiderhold 1997]. The Web is routinely used by students to access all kinds of information needed in classroom work or for research. When I taught a course at Columbia, the students complained about reading assigned on paper, much preferring the reading which was available on the Web. The Web is preferred not just because it has recreational content but also as a way of getting scholarly material.

The convenience of the Web is obvious. If I need a chart or quote from a Mellon Foundation report, I can bring it up in a few tens of seconds at most on my workstation. If I need to find it on paper, and it isn't in my office, I'm faced with a few minutes to visit the Bellcore library, and probably a few weeks since like most libraries they are cutting back on acquisitions and will have to borrow it from somewhere else. The Web is so convenient that I frequently use it even to read publications that I do have in my office.

Web use is greeted so enthusiastically that volunteers have been typing in (or scanning) out-of-copyright literature on a large scale, as for example for Project Gutenberg. The figure below shows the number of books *added* to the Project Gutenberg archive each year in the 1990s; by comparison in the entire 1980s only two books were entered.



By comparison, some of the electronic journal trials seem disappointing. Some of the reasons that digital library experiments have been less successful than they might have been involve the details of access. Whereas Web browsers are by now effectively universal on campuses, the specific software needed for the CORE project, as an example, was somewhat of a pain for users to install and use. Many of the electronic library projects involve scanned images which are difficult to manipulate on small screens, and they have rarely involved material which was designed for the kind of use that is common on computer systems. By contrast, most HTML material is written with the knowledge of the format in which it will be read and is adapted to that style. I note anecdotal complaints even that Acrobat documents as not as easy to read as normal Web pages.

Web pages, in particular, may have illustrations in color, and even animations, beyond the practical ability of any conventional publisher. Only one in a thousand pages of a chemical journal, for example, is likely to have a color illustration. Yet most popular web pages have color (although the blinking colored ad banners might be thought to detract rather than help Web users). Also, Web pages need not be written to the traditional standards of publishing \- the viewgraphs that represent the talk associated with a scholarly paper may be easier to read than the paper itself.

This suggests that the issue with the popularity of the Web compared with digital library experiments is not just content or convenience but also style. In the same way that *Scientific American* is easier to read than traditional professional journals, Web pages can be designed to be easier for students to read than the textbooks they buy now. Reasons might include the way material is broken into fairly short units, each of which is easy to grasp; the informal style; the power of easy cross-referencing, so that details need not be repeated, the extreme personality shown by some Web pages, and the use of illustrations as mentioned before. Perhaps some of these techniques, well known to professional writers, could be encouraged by universities for research writing.

The attractiveness of the newer Web material also suggests that older material will become less and less read. In the same way that vinyl records have suddenly become very old, or that TV stations refuse to show black-and-white movies, libraries may find that the 19th century material in many libraries disappears from the view of the students. Mere scanning to produce bitmaps, resulting in material which can not be searched and which does not look like newly written text, may produce material that although more accessible than the old volumes, is still not as welcome to students as new material. How much conversion of the older bitmaps can be justified? Of course many vinyl recordings are reissued on CD, and movies are colorized, but libraries are unlikely to have resources to do much updating. How will we be able to present the past in a way that students will be willing to use? Perhaps this will be a golden age for scholars as nearly the entire world supply of reference books will have to be rewritten for HTML.

Risks of the Web

Of course, access to Web pages typically does not involve the academic library or bookstore at all. What does this mean for the future of access to information at a university? There are threats to various traditional values of the academic system.

> **Quality.** Much of the material on the Web is junk; Gene Spafford refers to Usenet as a herd of elephants with diarrhea. Are students going to come to rely on this junk as real? Would we stop believing that slavery or the Holocaust really happened if enough followers of revisionist history put up a predominance of web pages claiming the reverse?
>
> **Loyalty.** It has already been a problem for universities that the typical faculty member in surface effect physics, for example, views his or her colleagues as the other experts in surface effect physics around the world, rather than the other members of the same physics department. Will the Web now mean that this is true of undergraduates as well? Will University of Michigan undergraduates read web pages from Ohio State? Can the Midwest survive that?
>
> **Shared experience.** Santayana wrote that it didn't matter what books students read as long as they all read the same thing. Will the great scattering of material on the Web mean that few undergraduate will be able to find somebody else who has been through the

same courses reading the same books? When I was an undergraduate I once had a friend who would look at people's bookshelves and recite the courses they had taken. This will become impossible.

Diversity. Since we can always fear two contradictory dangers, perhaps the ease of getting a few well-promoted Web site will mean that fewer sources are read. If nobody wants to waste time on a Web site that does not have cartoons, fancy color pictures and animation, then only a few well-funded organizations will be able to put up web sites that get an audience. Again, the United States publishes about 50,000 books each year, but produces less than 500 movies. Will the switch to the Web increase or decrease the variety of materials read at a campus?

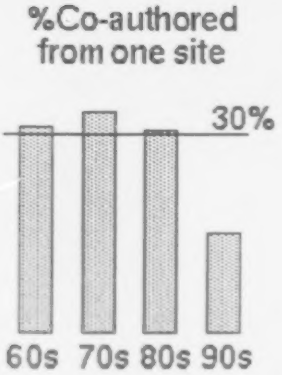
Equality of access If computers are needed to find information, will this produce barriers for people who lack money, good eyesight, or some kinds of interface-using skills? Universities want to be sure that all students can use whatever information delivery techniques are used; is the Web acceptable to at least as wide a span of students as the traditional library?

Recognition. Traditionally faculty obtain recognition and status from publishing in prestigious journals. High-energy physicists used to get their latest information from *Physical Review Letters;* today they rely on Ginsparg's preprint bulletin board at Los Alamos National Laboratory. Since this is not referred, how do people select what to read? Typically, they choose papers by authors they have heard of. So the effect of the switch to electronic publishing is that it is now harder for a new physicist to attract attention.

A broader view of threats posed by electronics to the university, not just those arising from digital library technology, has been presented by Eli Noam [Noam 1995]. Noam worries more about video tapes and remote teaching via television, and the possibility that commercial institutions might attempt to supplant universities, offering cheap education based entirely on electronic technologies. Should they succeed in attracting enough customers to force traditional universities to lower tuition costs, the financial structure of present-day higher education would be destroyed. Noam recommended that universities emphasize personal mentoring and one-to-one instruction to take the greatest advantage of physical presence.

Similarly, Van Alstyne and Brynjolfsson [Van Alstyne 1996] have warned of `balkanization' caused by the preference of individuals to select specialized contacts. They point to past triumphs involving cross-field work, such as the history of Watson and Crick, trained in physics and zoology respectively. In their view, search engines can be too effective, since letting people read only exactly what they were looking for may encourage overspecialization.

As an example of the tendency towards seeking collaborators away from one's base institution, the figure below shows the tendency of multi-authored papers to come from more than one institution. It was made by taking the first issue each year from the *SIAM Joural of Control and Optimization* (originally named *SIAM Journal of Control*) and counting the fraction of multi-authored papers in which all the authors came from one institution. The results was averaged over each decade. Note the drop in the 1990s. There has also, of course, been an increase in the total number of multiauthored papers (in 1965 the first issue had 14 papers and every paper had only one author; in 1996 there were 17 papers and only two were single-authored). But few of the multiple-authored papers today came from only one research institution.



Of course, there are advantages to the new technology as well, not just threats. And it is clear that the presence of the Web is coming, whatever universities do -- this is the first full paper I have written directly in HTML, rather than prepared for a typesetting language. Much of the expansiveness of the Web is all to the good; for many purposes access to random undergraduate opinions, and certainly to their fact-gathering, may well be preferable to ignorance. It is hard to imagine students or faculty giving up the speed with which things can be accessed from their desktops, anymore than we will give up cars because it is healthier to walk or ecologically more desirable to ride trains. How, then, can we ameliorate or prevent the possible dangers elaborated before?

University Publishing

Bellcore, like many corporations, has a formal policy for papers published under its name. These papers must be reviewed by management and others, reducing the chance that something sufficiently erroneous to be embarrassing, or something which poses a legal risk to the corporation, will appear. Many organizations do not yet have any equally organized policy for managing their web pages (Bellcore does have such a policy, dealing with an overlapping set of concerns). Should universities have rules about what can appear on their web pages? Should such rules distinguish between what goes out on `personal' or `organizational' pages? Should the presence of a page on a Harvard web page connote any particular sign of quality, similar to the appearance of a book under the Harvard University Press imprint? Perhaps a university should have an approved set of pages, providing some assurance of basic correctness, decency of content, and freedom from viruses; then people wishing to search for serious content might restrict their searches to these areas.

The creation of a university web site as the modern version of a university press or a journal offers a sudden switch back from publishers to the universities as the providers of information. If a university were to provide a refereed, high-prestige section of its web site, could it attract the publication that now goes to journals? The effect of this would be to provide a way for students to find quality material, and to build institutional loyalty and shared activities among the members of the university community. Perhaps the easiest way of doing this would be to make tenure depend on contribution to the university website, instead of contributions to journals.

The community could even be extended beyond the faculty. Undergraduate papers could be placed on a university web site; one can easily imagine different parts of the site for different genres ranging from the research monograph to the quip of the day. This would let all students participate and get recognition, so long as there is some quality control imposed on this part of the site and that presence on it is recognized as an honor.

In addition to supporting better quality, a university web site devoted to course reading could make sure that a diversity of views is supported. Online reading lists, just like paper reading lists, can be compiled to avoid the problem of everyone relying on the same few sites. This would help, for example, if many of the search engines start making money by charging people to be listed higher in the list of matches (a recurrent rumor, but perhaps an urban legend). It would also push students to look at sites which perhaps lack fancy graphics and animation.

One could even imagine that excessive reliance on a university web site could produce too much inbreeding. If we lost the publications that now provide general prestige in favor of university web sites, will it be possible for a professor at a less prestigious university to put an article on the Harvard or Stanford web site? If not, how will anyone ever move up? I do not perceive this as likely to be a problem anytime soon; the reverse (a total lack of organizational identification) is more likely.

It is likely that web sites of this sort would not include anonymous contributions. The net is somewhat overrun right now with untraceable postings that often contain annoying or inflammatory material, ranging from the merely boring commercial advertising to the deliberately outrageous political posting. Having a place which did not allow this kind of material might help to civilize the Web and make it more productive.

Information Location

Some professors already provide Web reading lists, corresponding to the traditional lists of paper material. The average Columbia course, for example, has 3000 pages of paper reading (with an occasional additional audiotape in language courses). The lack of quality on the Web means that it will become more important for faculty to provide guidance to undergraduates about what to read there.

More important, it will be necessary for faculty to teach the skill of looking purely at the text of a document and making a judgment as to its credibility. Much of our ability to evaluate a paper document is based on the credibility of the publisher. On the Web, students will have to judge by principles like those of paleography. What do we know, if anything, about the source? Is there a motive for deception? How does the wording of the document read -- credibly or excessively emotionally? Do facts that we can check elsewhere agree with other sources?

The library will also gain a new role. Universities should provide a training service for how to search the Web, and the library is the logical place to provide that. Partly this is a result of the training of librarians in search systems, which are rarely studied formally by any other groups. In addition, the librarians are the only hope to keep the alternative old information sources in front of students until most of them are converted, which will take a while.

The art of learning to retrieve information may also bring students together. I once asked a Columbia librarian whether the advent of computers and networks in the dormitory rooms was creating a generation of introverted nerds lacking social skills. She replied that it was the reverse. In the days of card catalogs students were rarely seen together; each person searched the cards alone. Now, she said, she frequently saw groups of two or three students at the OPAC terminals, one explaining to the others how to do something. Oh, I said, so you're improving the students social skills by providing poor human interface software. Not intentionally, she replied. Even with good software, however, there is still a place for students helping each other find things, and universities can try to encourage this.

Much has been written about the `information rich' vs. the `information poor' and the fear that once a machine costing several thousand dollars is needed to gain information, poor people will be placed at a still greater disadvantage in society than they are today. In the university context, money may not be the key issue, since many university libraries provide computers for general use. However, some people face non-financial barriers to the use of electronic systems. These may include limited eyesight or hearing (which of course also affect the use of conventional libraries). More important is perhaps the difficulty that some users may have with some kinds of interface design. This ranges from relatively straightforward issues such as color-blindness, to complex perceptual issues involving different kinds of interfaces and their demands on different individuals. So far we do not really know whether some users will have a need for something other than whatever becomes the standard information interface; in fact we do not know whether some university students in the past had particular difficulties learning card catalogues.

Libraries may also be a good place to teach aspects of collaboration and sharing that will grow out of references as hyperlinking replaces traditional citation. Students are going to use the Web to cooperate in writing papers as well as finding information for them. The ease of including (or pointing to) the work of others is likely to greatly expand the extent to which student work becomes collaborative. Learning how to do collaborative work effectively and fairly is an important skill students can acquire. In particular, the desire to make attractive multimedia works, which may need expertise in writing, drawing, and perhaps even composing music, will drive us to encourage cooperative work. Given the start of this effort with quoting references, the library may be a place to teach cooperative software.

Students could also be encouraged to help organize all the information on the local web site. Why should a student's web page prefer local resources? Perhaps because some kind of academic credit is created for doing that. University web sites, to remain useful, will require constant maintenance and updating. Who is going to do that? Realistically, studets

New Creativity

There is a wide rush of new presentation modes on the Web. We are going to see applets implementing animation, interactive games, and many other new kinds of presentation modes. The flowering of creativity in this should be encouraged. In the early days of television and of movies, the amount of equipment involved was beyond the resources of amateurs, and universities did not play a major role in their development. By contrast, universities are important in American theatre and classical music. The Web is also an area in which equipment is not really a limitation, and universities have a chance to play a role.

This represents a chance for the university art and music departments to join forces with the library. Just as the traditional tasks of preparing reading lists and scholarly articles can move onto a university web site, so can the new media. The advantage of doing this with the library is that we can actually save the beginnings of a new form of creativity. We lack the first email message; nobody understood that it was worth saving. Much of early film (perhaps half the movies made before 1950) no longer survives. 1950s television is mostly gone for lack of recording devices. In an earlier age, the Elizabetheans did not place a high value on saving their dramatic works; of the plays performed by the Admiral's Men (a competitor to Shakespeare's company) we have only 10% or 15% today. We have a chance not to make the same mistake with innovative Web page designs, providing that such pages are supported in some organized way, rather than on computers in individual student dorm rooms.

Recognizing software as a kind of scholarship is a change for the academic community. The National Science Foundation tends to say ``we don't pay for software, we pay for knowledge," drawing a sharp distincton between the two. Even computer science departments have sometimes said that you can't get a PhD for writing a program. The new kinds of creativity will need a new kind of university recognition. Will we have honorary web pages instead of honorary degrees? We need undergraduate course credit and tenure consideration for web pages.

Software and data are new kinds of intellectual output which are not traditionally considered creative. Traditionally, for example, the design of a map was considered copyrightable; the data on the map, although representing more of the work, were not considered design and not protectable. In the new university publishing model, data should be a first-class item, whose accumulation and collection is valuable and leads to reward.

Switching to honoring a web page rather than a paper does have consequences for style, as discussed above. Web pages also have no size constraints; in principle there is no reason why a gigabyte could not be published by an undergraduate. Universities will need to develop both tools and rules for summarizing and accessing very large items, as needed.

Conclusion

To preserve access to quality information while also preserving some sense of community in a university, the academic institutions should take a more active view of their web sites. By using the Web as a reward, and as a way of building links between people, universities could serve a social purpose as well as an information purpose. The ample space and low cost of Web publishing provide a way to extend the intellectual community of a university, and to make it more inclusive. This may encourage students and faculty to work together, maintaining a local bonding of the students. The goal is to use university web publishing, information searching mechanisms, and rewards for new kinds of creativity to build a new kind of university community.

REFERENCES:

[Egan 1991]. ``Hypertext for the Electronic Library? CORE sample results," by D. E. Egan, M. E. Lesk, R. D. Ketchum, C. C. Lochbaum, J. R. Remde, M. Littman, and T. K. Landauer, *Proc. Hypertext '91*, pages 299-312, San Antonio, Texas (15-18 Dec. 1991).

[Entlich 1997] ``Testing a Digital Library: User Response to the CORE Project," by Richard Entlich, Lorrin Garson, Michael Lesk, Lorraine Normore, Jan Olsen and Stuart Weibel, to appear 1997.

[Hoffman 1996] ``New Metrics for New Media, Toward the Development of Web Measurement Standards," by T. P. Novak and D. L. Hoffman, Project 2000 White Paper, available on the Web at http://www2000.ogsm.vanderbilt.edu/.

[Hunter 1996] *TULIP Final Report*, by M. Borghuis, H. Brinckman, A. Fischer, K. Hunter, E. van der Loo, R, Mors, P. Mostert and J. Zilstra, Elsevier Science Publishers, New York, 1996: ISBN 0-444-82540-1.

[Negroponte 1996] ``Caught Browsing Again," *Wired*, issue 4.05 (May 1996).

[Noam 1995] ``Electronics and the Dim Future of the University," by Eli M. Noam, *Science* vol. 270, no. 5234, p 247 (October 13, 1995).

[Samuel 1964] ``The Banishment of Paperwork," by A. L. Samuel, *New Scientist*, vol. 21, no. 380, 529-530 (27 Feb 1964).

[Van Alstyne 1996] ``Could the Internet Balkanize Science?" by Marshall Van Alstyne and Erik Brynjolfsson, *Science*, vol. 274, no. 5292, p. 1479-1480 (29 November 1996).

[Wiederhold 1997] Gio Wiederhold, private communication.

* * *

For additional information about the conference, or The Andrew W. Mellon Foundation's scholarly communication initiatives, please contact Richard Ekman. For additional information about ARL or this web site contact Patricia Brennan, ARL Program Officer at (202) 296-2296.

Return to Office of Scholarly Communication Home Page

ARL Home ARL Scholarly Communication and Technology Home Page

© Association of Research Libraries, Washington, DC

Web Design by Angelo F. Cruz
Maintained by ARL Web Administrator
Last Modified: August 20, 1997

Scholarly Communication and Technology

Conference Organized by The Andrew W. Mellon Foundation

at Emory University
April 24-25, 1997

Copyright © of the papers on this site are held by the individual authors or The Andrew W. Mellon Foundation. Permission is granted to reproduce and distribute copies of these works for nonprofit educational or library purposes, provided that the author, source, and copyright notice are included on each copy. For commercial use, please contact Richard Ekman at the The Andrew W. Mellon Foundation.

Session #4 Patterns of Usage

Online Books at Columbia: Measurement and Early Results on Use, Satisfaction, and Effect

Carol A. Mandel
Deputy University Librarian
Columbia University

and

Mary C. Summerfield
Coordinator, Online Books Project
Columbia University Libraries

and

Paul Kantor
Consultant

. . . .

ONLINE BOOKS AT COLUMBIA

MEASUREMENT AND EARLY RESULTS ON USE, SATISFACTION, AND EFFECT

1. 1.EXECUTIVE SUMMARY
2. 2.CONCEPTUAL FRAMEWORK
3. 3.THE ONLINE BOOKS COLLECTION
    1. 3.1Design Of the Online Books Collection
    2. 3.2Development of the Online Books Collection
        1. 3.2.1Purchased Texts
        2. 3.2.2Collaborating Publishers And Their Books
    3. 3.3User Access to the Collection
        1. 3.3.1Formats and Functionalities Over Time
        2. 3.3.2Who Can Use The Collection
        3. 3.3.3Access Paths to CWeb Books
        4. 3.3.4Publicity Campaign
4. 4.USE OF ONLINE BOOKS
    1. 4.1Methodology For Studying Use of And Reactions to Various Formats
        1. 4.1.1Measurement Plans
        2. 4.1.2Documentation Measures for Use of Online Books
        3. 4.1.3Documentation Measures for Reactions to Online Books
    2. 4.2Socio-Technical Environment
        1. 4.2.1External Socio-Technical Environment
            1. 4.2.1.1Media Coverage Of the Internet
            2. 4.2.1.2Personal Computer Specifications & Pricing Trends
            3. 4.2.1.3Household Computer Penetration & Internet Access
        2. 4.2.2Columbia Socio-Technical Environment
            1. 4.2.2.1Campus Infrastructure: February 1997
            2. 4.2.2.2Community Perceptions of Access To Computing Resources
            3. 4.2.2.3Community Use of Online Resources
    3. 4.3Findings On Use Of Books In Online Collection
        1. 4.3.1Reference Books
            1. 4.3.1.1Total Use Over Time
                1. 4.3.1.1.1Concise Columbia Electronic Encyclopedia
                2. 4.3.1.1.2Columbia Granger's World of Poetry*
                3. 4.3.1.1.3The Oxford English Dictionary
                4. 4.3.1.1.4Garland Reference Works
            2. 4.3.1.2Host Computers for Reference Book Use
        2. 4.3.2Monographic Books
            1. 4.3.2.1Total Use Over Time
                1. 4.3.2.1.1The Past Masters Collection
                2. 4.3.2.1.2Columbia University Press Monographs
                    1. 4.3.2.1.2.1Social Work Books
                    2. 4.3.2.1.2.2Other Works
                3. 4.3.2.1.3Oxford University Press Monographs
            2. 4.3.2.2Host Computers for Monographic Book Use
            3. 4.3.2.3Use By Day And Time -- All Types of Books
        3. 4.3.3Session Analysis for Use of CWeb Books

         1. 4.3.3.1Two Weeks' Sessions - All Online Books
         2. 4.3.3.2Sessions for Social Work Books
      4. 4.3.4Analysis of Unique Users Of Online Books
         1. 4.3.4.1Use and Users
         2. 4.3.4.2Use Concentration
            1. 4.3.4.2.1Reference Books
            2. 4.3.4.2.2Non-Reference Books
         3. 4.3.4.3User Cohorts
            1. 4.3.4.3.1Reference Books
               1. 4.3.4.3.1.1The OED
               2. 4.3.4.3.1.2Grangers Reference Works:
            2. 4.3.4.3.2Non-Reference Books
               1. 4.3.4.3.2.1Past Masters
               2. 4.3.4.3.2.2Columbia University Press
               3. 4.3.4.3.2.3Oxford University Press
         4. 4.3.4.4Online Book Use Per User
5. 5.UNDERSTANDING FACTORS INFLUENCING USE
   1. 5.1Resource Factors
   2. 5.2Attitude Factors - In-Class Survey
      1. 5.2.1The In-Class Survey
      2. 5.2.2Methods of Studying A Class Reading
      3. 5.2.3Preferences for Studying Class Reading
      4. 5.2.4Reasons for Preference
      5. 5.2.5Impact of Preferred Method
      6. 5.2.6Comparison of Personal Behavior and Perception of Others' Behavior
6. 6.THE ONLINE SURVEY
   1. 6.1The Online Interview Instrument
   2. 6.2Early Results from CWeb Survey
      1. 6.2.1CWeb Survey Responses by Online Text Used
      2. 6.2.2CWeb Survey: Primary Project for Using Book
      3. 6.2.3CWeb Survey: Ways of Using Book
      4. 6.2.4CWeb Survey: Forms of This Book Ever Used
      5. 6.2.5CWeb Survey: Preferred Form of This Book
      6. 6.2.6CWeb Survey: Reasons for Preference
      7. 6.2.7CWeb Survey: Preferred Format and Reasons for Preference
      8. 6.2.8CWeb Survey: Frequency of Use in Past Three Months
      9. 6.2.9CWeb Survey: Total Usage in Minutes in Past Three Months
      10. 6.2.10CWeb Survey: Frequency of Use of Any Online Book in Past Year
      11. 6.2.11CWeb Survey: Effect of Online Books on Scholarly Work
          1. 6.2.11.1CWeb Online Survey: Productivity, Book Type and Format
          2. 6.2.11.2CWeb Online Survey: Work Quality, Book Format and Type
      12. 6.2.12CWeb Online Survey: Columbia Cohort of Respondents
      13. 6.2.13CWeb Online Survey: Discipline of Respondents
      14. 6.2.14Online Survey: Place in Project
   3. 6.3User Comments
7. 7.BEHAVIORAL FACTORS
   1. 7.1Access to Networked Computer
   2. 7.2Time In Online Activities
      1. 7.2.1Responses to CWeb Online Survey
      2. 7.2.2Responses to Questionnaire with Paper OED

3. 7.2.3Responses to In-Class Questionnaire, Fall 1996 and Spring 1997
8. 8.CONCLUSIONS

Figure 1. Interrelation of Factors Involved in the Use and Impact of Online Books.

Exhibit 1. Columbia University Digital Library Collections

Exhibit 2. Online Books Evaluation Project: Titles Included - Home Page

Exhibit 3. Online Books Evaluation Project: Titles Included - Current Social Sciences

Table 1. Types of Surveys

Table 2. New York Times Stories Involving Information Services

Table 3. Minimum Recommended Specifications for Home Computers, May 1994 - April 1997

Table 4. Characteristics of a $2,000 Computer, December 1994 - June 1997

Table 5. U.S. Internet Households

Table 6. March 1997 In-Library Survey: Is there a computer (in the library or elsewhere) attached to the campus network (directly or by modem) that you can use whenever you want?

Table 7. Fall 1995 Campus Survey: Is there a computer (in the library or elsewhere) attached to the campus network (directly or by modem) that you can use whenever you want?

Table 8. March 1997 In-Library Survey: Weekly Hours on Online Activities by Access to Computers Linked to Campus Network, Winter 1997

Table 9. March 1997 In-Library Survey: Weekly Hours In Online Activities by Columbia Status, Winter 1997

Graph 1. Concise Columbia Electronic Encyclopedia Sessions, 1994 - 1997: CNet

Table 10. Columbia Granger's World of Poetry: Number of Hits by Month

Table 11. Oxford English Dictionary: Number of Hits by Month

Table 12. Garland Reference Works: Number of Hits by Month, December 1, 1996 - May 31, 1997

Table 13. Host Computers for Reference Book Use, May 1, 1996 - March 15, 1997 - Percent Distribution

Table 14a. Past Masters On The Web, Total Monthly Hits: May 1995 - March 1997

Table 14b. Key Past Masters Texts On The Web, Monthly Hits: August 1996 - May 1997

Table 15. Scholarly Hits on Columbia University Press Books, May 1, 1996 - May 31, 1997

Table 16. Columbia Circulation of Columbia University Press Monographs: 1994 - 1996

Table 17a. Scholarly Uses (Hits) Oxford University Press Monographs In Literary Criticism, May 1, 1996 - May 31, 1997

Table 17b. Scholarly Uses (Hits) Oxford University Press Monographs In Philosophy, May 1, 1996 - May 31, 1997

Table 18. Columbia Circulation of Oxford University Press Monographs: 1995 - 1996

Table 19. Host Computers for Monographic Book Use, May 1, 1996 - May 31, 1997

Table 20. Patterns of Use from Web Server: May 1, 1996 - March 15, 1997 Hits

Table 21. Hits by Time of Day: May 1, 1996 - March 15, 1997

Table 22. Online Book Usage (Web): Hits Per Use Session

Table 23. Session Analysis for Social Work Books, Fall 1996

Table 24. Hits by Book Category, March 15 - May 31, 1997

Table 25. Unique Users and Hits for Reference Books, March 15 - May 31, 1997

Table 26. Non-Reference Books by Number of Users, March 15 - May 31, 1997

Table 27. Unique Users and Hits for Non-Reference Books, March 15 - May 31, 1997

Table 28. Distribution of Hits Per Unique User, March 15 - May 31, 1997

Table 29. Distribution of Unique Titles Viewed Per User, March 15 - May 31, 1997

Table 30. Hits Per Unique User by Academic Cohort, March 15 - May 31, 1997

Exhibit 4. Survey Of Book Use For Course Readings

Table 31. Methods of Reading This Assignment: Whole Sample Fall 1996 And Spring 1997

Table 32. Preferred Method Reading This Assignment: Whole Sample Fall 1996 and Spring 1997

Table 33. Reason for Preferred Method: Whole Sample Fall 1996 and Spring 1997

Table 34. Preferred Method & Reason for Preferred Method: Whole Sample Spring 1997 - Row Percentages

Table 35. Nature of Impact of Preferred Method: Whole Sample Fall 1996 and Spring 1997

Table 36. In-Class Surveys: Personal Behavior and Perception of Methods Used by Classmates to Read This Assignment: Whole Sample Fall 1996 and Spring 1997

Exhibit 5. Online Survey Instrument: Non-Reference Books

Table 37. CWeb Online Survey Responses by Online Text Used, September 1996 - June 1997

Table 38. CWeb Online Survey Responses by Work Involved, September 1996 - June 1997

Table 39. CWeb Online Survey Responses: Uses of The OED , September 1996 - June 1997

Table 40. CWeb Online Survey Responses: Uses of Other Books, September 1996 - June 1997

Table 41. CWeb Online Survey Responses: Forms of the Book Ever Used, September 1996 - June 1997

Table 42. CWeb Online Survey Responses: Preferred Book Form, September 1996 - June 1997

Table 43. CWeb Online Survey Responses: Reasons for Book Form Preference, September 1996 - June 1997

Table 44. CWeb Online Survey Responses: Preferred Book Form and Key Reasons for Preference, September 1996 - June 1997

Table 45. CWeb Online Survey Responses: Frequency of Use of This Book in Past three Months, September 1996 - June 1997

Table 46. CWeb Online Survey Responses: Total Usage of This Book In Minutes in Past Three Months, September 1996 - June 1997

Table 47. CWeb Online Survey Responses: Total Usage of Online Books In Past Year, September 1996 - June 1997

Table 48. CWeb Online Survey: In doing the type of work for which you used this book, do paper books or online books help you be more productive? by Book, September 1996 - June 1997

Table 49. CWeb Online Survey: Do you find that you are able to do work of higher quality when you use paper books or online books? by Book, September 1996 - June 1997

Table 50. CWeb Online Survey: Quality and Productivity, September 1996 - June 1997

Table 51. CWeb Online Survey: Respondent's Columbia Status, September 1996 - June 1997

Table 52. CWeb Online Survey: Respondent's Discipline, September 1996 - June 1997

Exhibit 6. An Extended Comment

Table 53. In-Class Surveys: Ready Access to Networked Computer: Whole Sample Spring 1997

Table 54. In-Class Survey: Preferred Method of Reading This Assignment and Access to Networked Computer: Whole Sample Spring 1997

Table 55. CWeb Online Survey: Weekly Hours In Online Activities, September 1996 - June 1997

Table 56. Weekly Hours In Online Activities for 11 Respondents to Paper Questionnaire on Use of OED, 1996 -- Percent of Respondents

Table 57. In-Class Surveys: Weekly Hours In Online Activities: Whole Sample Fall 1996 and Spring 1997

Table 58. In-Class Surveys: Mean Weekly Hours In Online Activities By Class Type: Fall 1996 and Spring 1997

Table 59. In-Class Surveys: Weekly Hours In Online Activities By Access to Networked Computer: Spring 1997

Appendix 1. Contemporary Books In the Online Collection

Appendix 2. Format Availability, Winter 1997, and Introduction Date

Appendix 3. Home Computers: Minimum Specifications Recommended

Appendix 4. Gateway 2000 Home Computers: System Specifications

Appendix 5. U.S. Household PC & Internet Penetration

Appendix 6. Concise Columbia Electronic Encyclopedia Sessions, 1994 - 1997: CNet

....

Cover Page Table of Contents Next Page Footnotes

Previous Page

For additional information about the conference, or The Andrew W. Mellon Foundation's scholarly communication initiatives, please contact Richard Ekman. For additional information about ARL or this web site contact Patricia Brennan, ARL Program Officer at (202) 296-2296.

Return to Office of Scholarly Communication Home Page

Scholarly Communication and Technology

Conference Organized by The Andrew W. Mellon Foundation

at Emory University
April 24-25, 1997

Copyright © of the papers on this site are held by the individual authors or The Andrew W. Mellon Foundation. Permission is granted to reproduce and distribute copies of these works for nonprofit educational or library purposes, provided that the author, source, and copyright notice are included on each copy. For commercial use, please contact Richard Ekman at the The Andrew W. Mellon Foundation.

Session #4 Patterns of Usage

Online Books at Columbia: Measurement and Early Results on Use, Satisfaction, and Effect

Carol A. Mandel
Deputy University Librarian
Columbia University

and

Mary C. Summerfield
Coordinator, Online Books Project
Columbia University Libraries

and

Paul Kantor
Consultant

* * *

1. EXECUTIVE SUMMARY

The Online Books Evaluation Project at Columbia University explores the potential for online books to become significant resources in academic libraries by analyzing (1) the Columbia community's adoption of and reaction to various online books and delivery system features provided by the Libraries over the period of the Project; (2) the relative life cycle costs of producing, owning and using online books and their print counterparts; and (3) the implications of intellectual property regulations and traditions of scholarly communications and publishing for the online format.

Online books might enhance the scholarly processes of research, dissemination of findings, teaching, and learning. Alternatively, or in addition, they might enable publishers, libraries and scholars to reduce the costs of disseminating and using scholarship. For example:

- If the scholarly community were prepared to use some or all categories of books for some or all purposes in an online format instead of a print format, publishers, libraries and bookstores might be able to trim costs as well as enhance access to these books.
- If online books made scholars significantly more efficient or effective in their work of research, teaching and learning so as to enhance revenues or reduce operating costs for institutions of scholarship, these books might be worth adopting even if their costs were no lower than those for their print counterparts.
- If an online format became standard, publishers might be able to offer affordable online access to books to institutions which would not normally have purchased print copies, thus both expanding convenient access to scholarship to members of those institutions and expanding publishers' revenues from these books.

The Columbia Online Books Evaluation Project is designed to learn about the scholarly community's enthusiasm for the online format in the near term and about features that users will demand, to project likely adoption patterns, and to estimate gains in operating effectiveness, revenue, and cost, if any, to be realized by publishers, funders of scholarship, libraries, and scholars. The Project confronts and explores a set of feasibility issues, including publishers' ability to provide books of various types and vintages in forms conducive to conversion to online formats and our ability to convert them to online books that will serve users' needs and preferences.

This paper focuses on the first of the Project's elements, user response, and reports on:

1. the conceptual framework for the Project; (Section 2)
2. background information on the status of the collection and other relevant Project elements; particularly design considerations; (Section 3)
3. our methodology for measuring adoption of online books by the Columbia community; (Section 4.1)
4. our current findings on relevant environmental factors, including access to online resources; (Section 4.2)
5. our current findings on use of online books and other online resources; (Section 4.3, 5.2)

6. our current findings on attitudes toward online books. (Sections 6 and 7)

The paper also reflects on our experience as a case study, specifically (1) problems encountered evaluating online resources (Section 6), and (2) problems encountered in producing online books (Section 3.2). The Project began in early 1995 with three reference works online; in autumn 1996 the first modern monographs became available to the Columbia University community which is the focus of the study.

Our current findings may be summarized as follows:

- *Most if not all reference books are used more heavily online than in print.* (Section 4.3.1)
- *Early online reference books have experienced falling usage over time, substitution of use of a new delivery system for an old one, or a smaller rate of growth of use than might be expected given the explosion in access to and use of online resources in general.* (Section 4.3.1)

In the early to mid-1990s, the novelty of online books may have brought users to the format somewhat without concern for their design, the utility of the delivery system, or the qualities of the books. With enhancement in delivery systems and expansion in the number of online books, being online is no longer a guarantee that a book will be used. New graphical delivery systems offer superior performance that is likely to draw scholars away from these early online resources provided via text-based systems increasingly, as access to those new delivery systems spreads. In addition, as more competing resources come online and either provide information that serves the immediate needs of a user better or offers a more attractive, user friendly format, scholars are less likely to find or to choose to use any single resource.

- *Online scholarly monographs are available to and used by more people than their print counterparts in the library collection.* (Section 4.3.2) Once a print book is in circulation, it is effectively unavailable to others for hours in the Reserve collection and weeks or months in the regular collection. An online book is always available to any potential user who has access to a computer with a Web browser.
- *Being online may bring to a book scholars who would not have seen it otherwise.* (Section 4.3.2) However, it is not yet clear whether their productivity or work quality will be significantly enhanced by such serendipity. The important concept of collation is transformed, in the networked environment, to a diversity of finding and navigational systems. As the online collection grows, browsing will require the focused use of online search tools rather than use of project-oriented Web pages.
- *Data from the most recent 11 weeks, including the last half of the spring 1997 semester, suggest that when a social work book available in both print and online formats was used in a course, the share of students using the online version was at most one-quarter.* (Section 4.3.4) We will track this rate of penetration for social work and other disciplines over the next semesters to see if students increase their rate of adoption.
- *Some scholars, especially students in a course assigned a reading that is in the online collection, are looking at the online books in some depth, suggesting that they find value in this means of access.* (Section 4.3.4) For example, in the most recent 11 weeks

analyzed, the two most frequently used monographs averaged 9.6 and 7.7 hits (chapters) per unique user.

- *Scholars residing off-campus are not using the online books from their homes to a significant degree.* (Section 4.3.2.2) For the ten months from May 1996, only 11 percent of the hits on Columbia University Press monographs were dial-up connections. Scholars report (in our interviews) that the expense of dialing-in to campus or maintaining an Internet account, the lack of sufficiently powerful home computers and Web software, and the slowness of delivery of the Web over standard modems are key constraining factors.

- *Students residing on campus may have Ethernet connections to the campus network - providing both speedy and virtually free access to the online collection. They are using online books, especially reference works, from their dorms during hours when the libraries are not open.* Forty-two percent of the hits on *The Oxford English Dictionary* in the ten months from May 1996 were from computers with such residence hall connections. (Section 4.3.1.2)

- *Some scholars perceive gains in the productivity and quality of their work in using online books, particularly reference books. Over half the respondents to our online survey (a small number) see the productivity and quality of their work using online resources to be as good or better than that achieved using paper resources.* (Section 6.2.11)

- *In surveys and interviews, students report that they particularly value easy access to the texts that are assigned for class and an ability to underline and annotate those texts. Students seek the ability to print out all or parts of the online texts that they are using for their courses, again indicating their desire to have the paper copy to use in their studying. Computer access to a needed text is not equivalent to having a paper copy (whole book or assigned portion) in one's backpack, available at any time and at any place.* (Section 5.2.4)

If the effective choice is between borrowing a book from the library, probably on a very short term basis from Reserves, and accessing the book online, the student is facing a parallel situation of needing to photocopy or print out to obtain portable, annotatable media. However, the online book has the advantages of never being checked out when one wants to use it and of being accessible from a computer anywhere in the world at any time (as long as that computer has an Internet connection and a browser).

- *In surveys and interviews we find that scholars value the ability to do searches, to browse, and to look up information in an online book quickly. They also like the ability to clip bits of the text and put them in an electronic research notes file. Willingness to browse and to read online for extended periods varies from person to person, but it does not seem to be widespread at this time.*

- *Scholars with easy access to a networked computer spend more time online and are more likely to prefer to use one of the forms of the online book.* (Section 7) This suggests that, over time as such access achieves greater penetration in the scholarly community, online books will be achieve greater acceptance.

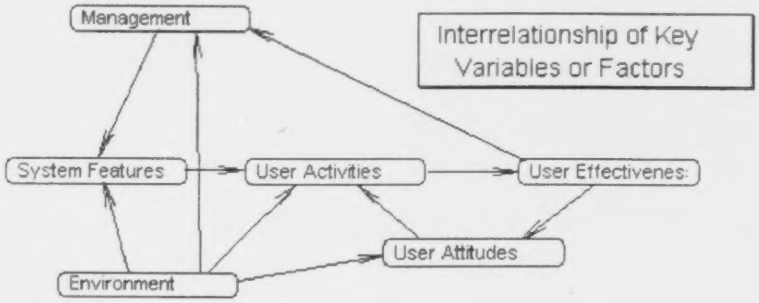
- *In the most recent 11 weeks studied, 52 percent of the online book users who viewed our online survey responded to it, but only 15 percent of these users chose to click on the button taking them to the survey.* (Section 6.2.14) Designing an online survey that is available to the reader without his taking action might enhance the response rate significantly.

2. CONCEPTUAL FRAMEWORK

The variables representing usage of a system of scholarly communication and research are at the same time effects and causes. Since scholars, the users of the system, are highly intelligent and adaptive, the effect of the system will influence their behavior, establishing a kind of feedback loop. As the diagram in Figure 1 shows, there are two key loops. The upper one, shown by the dark arrows, reflects an idealized picture of university administration. In this picture, the features of any system are adjusted so that, when used by faculty and students, they improve institutional effectiveness. This occurs in the context of continual adaptation on the part of the users of the system, as shown by the lighter colored arrows in the lower feedback loop.

All of this is constrained by the continual change of the environment, which affects the expectations and activities of the users, affects the kind of features that can be built into the system, and affects the very management that is bringing the system into existence. This interaction is shown by the dotted arrows in Figure 1.

Figure 1. Interrelation of Factors Involved in the Use and Impact of Online Books.



Our primary research goal, in relation to users, uses, and impacts, is to understand these relationships, using data gathered by library circulation systems, Internet servers, and surveys and interviews of users themselves.

. . . .

Cover Page Table of Contents Next Page Footnotes

Previous Page

Scholarly Communication and Technology

Conference Organized by The Andrew W. Mellon Foundation

at Emory University
April 24-25, 1997

Copyright © of the papers on this site are held by the individual authors or The Andrew W. Mellon Foundation. Permission is granted to reproduce and distribute copies of these works for nonprofit educational or library purposes, provided that the author, source, and copyright notice are included on each copy. For commercial use, please contact Richard Ekman at the The Andrew W. Mellon Foundation.

Session #4 Patterns of Usage

Online Books at Columbia: Measurement and Early Results on Use, Satisfaction, and Effect

Carol A. Mandel
Deputy University Librarian
Columbia University

and

Mary C. Summerfield
Coordinator, Online Books Project
Columbia University Libraries

and

Paul Kantor
Consultant

* * *

3. THE ONLINE BOOKS COLLECTION

The Online Books Evaluation Project began formal activity in January 1995. However, discussions with publishers about cooperating in such an effort by providing books and collaborating in research began in 1993, if not earlier. As noted in the Project's *Analytical Principles and Design* document, "The Online Books Evaluation Project is a component of the developing digital library at Columbia University. As part of its digital library effort, the Columbia University Libraries is acquiring a variety of reference and monographic books in electronic format to be included on the campus network; in most cases, those books will be available only to members of the Columbia community. Some of the books are being purchased; others are being provided on a pilot project basis by publishers who are seeking to understand how the academic community will use online books if they become more widely available in the future."

Columbia University Libraries provides the Columbia community with access to a substantial and growing set of full text (journals and reference materials), image, data and bibliographic online resources in addition to those that we are studying in the Online Books Evaluation Project. Some have been acquired or developed at Columbia and are maintained on servers here, e.g., art images, working papers. Others are maintained by publishers with access licensed to Columbia, e.g., *Encyclopedia Britannica* and Gale's *Contemporary Authors* and *Encyclopedia of Associations*. Yet others are maintained elsewhere and access is free to all, with Columbia subject specialists providing links on their subject home pages.

3.1 Design Of the Online Books Collection

When this Project was proposed, the World Wide Web was an emerging technology, and we still expected to develop specialized browsers for using the books in SGML format, just as other online projects were doing at the time. However, by the time the Project was funded and ready to mount books online, it was clear that the Web would soon be the best delivery system for maximizing availability of the books to the scholarly community. Web browsers had, and still have, annoying limitations, but we felt that they would become better over time and provide optimum flexibility to users.

Many other online projects are providing users with materials in PDF, scanned, or bitmapped format. These are effective formats for journal articles, which are finely indexed through existing sources and which are short and easily printed. However, the greatest potential added value from online books, compared to their print counterparts, comes with truly digital books. Only in this type of format, for example, can users do full search for terms or cut and paste parts of the book to another document. In addition, only this online format allows the development of truly interactive books that take advantage of the current and anticipated capabilities of Web technology, such as the inclusion of sound and video, data files and software for manipulating data, and links to other online resources. Perhaps only such enhanced online books will offer sufficient advantages over traditional print format that scholars will be willing to substitute them for the print format for any or all of their modes of use and for any or all classes of books.

We have devoted considerable time and effort over the past two years to dealing with technical and design issues for the books. The design has evolved over this period as Web technology has advanced and as the Project team and users have reacted to early decisions. We will continue to work with users over the months ahead in order to provide basic design features that they endorse. We hope to begin to introduce more interactive features as appropriate to various books and to measure user response to them.

We look forward to comparing the results of our evaluations with those of online projects using other formats to explore whether format does make a significant difference in user attitudes and behavior.

3.2 Development of the Online Books Collection

3.2.1 Purchased Texts

Purchased texts included in the Online Books collection are *The Oxford English Dictionary* and classical texts in social thought from InteLex's *Past Masters* CD-ROM. Columbia converted the *Past Masters* texts from SGML to HTML for Web access. Ten *Past Masters* texts were made available to the Columbia community online in mid-1995, although with little publicity. Another 44 went online in July 1996, with publicity for the collection beginning early in the Fall. We intend to convert several other purchased CD-ROM products, largely literary texts, and include them in the collection in the near future. The Columbia digital library provides access to many other full text works to the scholarly community, but the ones described here have been the focus of our analysis, in large part because they are mounted on local servers from which detailed usage information can be gathered.

3.2.2 Collaborating Publishers And Their Books

Publishers participating in the Project by providing electronic files for their books and collaborating in the research effort are Columbia University Press, Garland Publishing, Oxford University Press, and Simon and Schuster Higher Education. All but Garland have been involved in this Project since its inception; Garland jointed the effort in 1996. The books provided by each publisher and the timing of the introduction of those books to the online collection are as follows:

Columbia University Press: Two reference works, *The Columbia Granger's Index to Poetry* and *The Concise Columbia Electronic Encyclopedia*, have been available since the outset. Columbia will provide three more reference books - *The Columbia Electronic Encyclopedia, The Columbia Guide to Standard American Usage*, and *The Columbia World of Quotations* - in 1997. Monographs, anthologies and textbooks are being provided in the fields of social work, literary criticism, political science, and earth and environmental science. The Project includes only books for which the Press can obtain both electronic files and author permissions. Sixteen such books are now in the collection, seven of them in the field of social work. The first of these books were made available online in September 1996. At this point, it appears that 27 more CUP books published in these fields in the past three years will be available to our collection; they will be added in the next few months.

Garland Publishing: Three Garland reference works, *The Chaucer Name Dictionary, Native American Women: A Biographical Dictionary*, and *African American Women: A Biographical Dictionary*, were added to the collection from December 1996 through February 1997. We selected these books because Columbia has sizable user groups in Medieval and Women's Studies and because they were available in electronic format and amenable to conversion to HTML. Garland is reviewing its collection and its resource availability to determine whether it can provide any other books to the Project.

Oxford University Press: In 1995, Oxford agreed to provide its monographs in the fields of literary criticism, neuroscience, and philosophy from the publication lists for 1995 through 1997. Oxford reports that a substantial share of titles in these fields have low sales and, hence, represent *the endangered scholarly monograph.* As of early 1996, Oxford had provided electronic files for 19 monographs in the fields of literary criticism and philosophy. Oxford required the Project to provide an online ordering mechanism concurrent with the availability of its books; that ordering system was ready for use in October. Sixteen Oxford books were online by year end 1996; 17 are now online. In June 1997, Oxford provided nine more books in literary criticism and philosophy. These should be online by fall 1997.

Simon and Schuster Higher Education: By late 1994, Simon and Schuster had agreed to contribute high use titles, defined as books on reserve for Columbia courses that had relatively heavy circulation. Simon and Schuster provided electronic files for nine such books, most of them in business-related subjects, in Fall 1995. As of June 1997, two of the books were online and the others were expected to be ready before the new academic year.

3.2.3 The Challenge of Obtaining Electronic Files for Books from Publishers

The Project's 1997 Annual Report discusses publishers' difficulty in providing electronic files for books that are amenable to conversion to the HTML format being used in the Project. Those problems include:

- Neither the publishers nor their printers have ready access to the final electronic files, e.g., typesetter's tapes, for books unless specific provision has been made for systematic retention and archiving of such files. Most publishers have not been able routinely to provide the Project with copies of the electronic files for books published since the early to mid-1990s.
- The electronic files for some books contain so many special characters and graphics that conversion to HTML format is infeasible.
- Publishers never possessed electronic files for books that authors supplied as camera-ready copy.
- After publication, seeking permissions from multiple copyright owners involved in a book, such as a collection of essays, would be too onerous.
- Interviews with authors reveal that those who refuse to include their books in the Project do so for various reasons. Some fear the ease of downloading and printing Web materials will tempt users not to respect copyright and that scholars outside of the Columbia community will receive copies of their works, thus reducing their royalty income. Others oppose the concept of online books and do not want to encourage them.

3.3 User Access to the Collection

3.3.1 Formats and Functionalities Over Time

As of June 1997, the Columbia community had access to a total of 96 online texts that are part of the Online Books Project. The Libraries have each book in print form, circulating from the regular collection or Reserves, or non-circulating in Reference, as well as in one or more online formats. Appendix 1 summarizes the print access modes for all the modern books in the collection. The various online modes have differing functionalities beyond browsing or reading on screen. Appendix 2 summarizes the schedule of mounting for the online books and their functionalities.

3.3.2 Who Can Use The Collection

By agreement with the publishers, we restrict access to the Project's online books to members of the Columbia University community, i.e., faculty, staff and students of Columbia and affiliated institutions who use the books in the Libraries and from anywhere via network access. Until March 1997, books were also available, only on Libraries terminals, to alumni and others with reading privileges. This policy both protects the publishers' intellectual property and provides the Project with the ability to gather richer data on usage.

Through Winter 1997: We employed two methods through Winter 1997 to maintain this control of access.

- To use books on CNet or at the Unix prompt, a scholar must sign in with her Columbia email address and password. This remains the case for this set of books. The exception to this rule is the public CNet terminals.
- To access books on CWeb, a scholar was required to use a computer with an address that the server recognizes as Columbia affiliated. Members of the Columbia community who connected to CWeb from a service like AOL were not able to use the collection. On the other hand, guests using X-terminals on the Columbia campus could reach those books.

In both cases, the data the server logs did not include information on the user. We initially planned to develop a directory of Columbia IP addresses by location and to link it to the server data in order to make general discrimination between dormitories and various other campus buildings. However, we decided that developing and maintaining this database would be too costly, given our near term plans for individual user authentication. Instead our analyses for the period before mid-March 1997 result from deduction based on the host name of the user computer.

As of March 15, 1997: For books in this Project and other materials with user restrictions, Columbia has developed and deployed a more robust system for Web authentication and access. This system permits a member of the Columbia community to use materials even if she connects through an Internet service provider like AOL. It requires each user to sign in when he wants to use one or more items in the collection. During a session, he needs to sign in only once. Ultimately, data records will be session based, that is linking all the activities by a user in a single session within its umbrella into a single record and providing information on the identity of that user.

Future: Given that Web browser/server interaction are *stateless*, i.e., each transaction is essentially independent of previous ones and the server retains no memory of a user's previous actions, translating the ability to control access to resources to the Web has been a challenge.

This local authorization system manages access with information from the central authentication database. This session-based system supports more extensive analysis of usage patterns. In particular, usage statistics can be tied to user characteristics. The management statistics system that will link access to a book with information on the user's affiliations and status should be fully ready this summer. In particular, transaction statistics may be aggregated for individuals, based on their initial 'login', providing more continuous, 'session based' tracking. To protect privacy, the personal key will be retained long enough to look up the required demographic information and will then be retained in encrypted form, to serve as an anonymous unique identification code.

3.3.3 Access Paths to CWeb Books

Users have six main alternatives for learning of the CWeb books: (1) word of mouth; : (2) the online catalog; (3) the Libraries' Digital Collections Web page; (4) the Project's home page; (5) Web pages for specialized library collections; and (6) publicity flyers, email messages, and formal and informal presentations by librarians and Project staff directed at the faculty and students most likely to be interested in the various online book collections.

In CLIO, the online catalog, a record for each online book lists its Web address (URL). In the near future when CLIO moves to the Web, a scholar will be able to click on that URL in the CLIO record and proceed directly to the book. During the period covered by this report, however, in order to move from that CLIO record to the online book, the scholar must either copy or write out the URL, switch to the Web, and input the URL into the Location box.

The first CWeb access point for the monographic (non-reference) books is a set of links to the Web pages with the subject categories into which we have grouped the books and another link to an alphabetical listing by author of all the texts in the collection. These links are on the Libraries' Digital Collections home page at http://www.columbia.edu/cu/libraries/digital/. (See Exhibit 1.)

Exhibit 1. Columbia University Digital Library Collections

http://www.columbia.edu/cu/libraries/digital/

A scholar starting at the Columbia University Web home page must take two steps to reach that list (to *Libraries*, to *Digital Collections*).

During Fall 1996 and Winter 1997, we sought ways to focus user attention on the collection, in the hopes of achieving more use and feedback. At the end of 1996, we launched a new Project home page (http://www.columbia.edu/dlc/olb/); see Exhibit 2. This page has a brief description of the evaluation effort, a link to the page that includes copies of the Project documents, a button for comments about the design of the online books system, a button for sending email to the Project Coordinator, and a capability to search by keyword throughout the books in the collection. In addition, it has links to groups of books in the collection: *Historical Social Thought, Current Humanities, Current Social Science, Current Science,* and *Current Reference.* We have included books in more than one of those groupings as appropriate; for example, each Garland reference book is in *Current Reference* and another subject category.

Once the scholar moves to one of the topical collection pages, he sees the books arrayed by primary subject category; pictures of the books' dust jackets accompany some of the titles. (Exhibit 3 has part of the *Current Social Science* page; http://www.columbia.edu/cu/libraries/digital/texts/social_sciences.html.) He has two options at this point: (1) clicking on one of those titles and going directly to the Table of Contents for that books or (2) doing a keyword search on that whole topical collection.

Besides these core locations, the online books on CWeb are typically linked to several pages where potential users might find them. Most of these are subject listings that collection bibliographers maintain, e.g., *Online Books* on the Social Work Library home page links to the *Current Social Science* page, or the Medieval Studies home page listing of Internet resources links to *The Chaucer Name Dictionary*.

A scholar wishing use one of the online collections repeatedly could bookmark the relevant subject matter page. He would then need only to select that bookmark from within his browser in order to reach that page.

The five Web reference books in the Online Books collection are also included in a separate set of pages maintained by the Reference Department. The scholar must traverse several levels before reaching any of the resources using this route. Finally, some of these resources are linked to Web pages created by various other Columbia groups.

Exhibit 2. Online Books Evaluation Project: Titles Included - Home Page

http://www.columbia.edu/dlc/olb/

Exhibit 3. Online Books Evaluation Project: Titles Included - Current Social Sciences

http://www.columbia.edu/cu/libraries/digital/texts/social_sciences.html

3.3.4 Publicity Campaign

Our publicity campaign for the online books collection has had several facets. The key component is a set of flyers, each focusing on one category of books. These flyers have major headlines followed by a listing of the online books available in that category, a brief explanation of the Online Books Evaluation Project, and then directions on how to reach and use the collection. These flyers have been sent to all the faculty members in each of the related departments and to graduate students whom we have identified as teaching in those departments. In some cases in which faculty members are using one of the titles in a course, we have provided copies of the flyer to each student. In some cases, we have gone to those classes to discuss the Project and how to use the books. We have also made presentations to faculty groups about the Project. More such presentations will be made in future semesters.

At this point, we are seeking a viable balance in our publicity. Over-promoting a collection that contains only a few books may create disgruntled potential users who are likely to be skeptical about the collection in the future. On the other hand, publicity is needed in order to create the awareness and sampling that are necessary precedents to regular use of online materials. Marketing research shows that publicity is most successful in cases in which a target group is generally seeking the product being offered. In our case that is scholars are likely to focus on publicity when they need to use one or more of the available books, e.g., Social Work students who are using one of the titles in a course or undergraduate students who have been told to use *The OED* for an assignment.

……

Cover Page **Table of Contents** **Next Page** **Footnotes**

Previous Page

For additional information about the conference, or The Andrew W. Mellon Foundation's scholarly communication initiatives, please contact Richard Ekman. For additional information about ARL or this web site contact Patricia Brennan, ARL Program Officer at (202) 296-2296.

Return to Office of Scholarly Communication Home Page

ARL Home ARL Scholarly Communication and Technology Home Page

© Association of Research Libraries, Washington, DC

Web Design by Angelo F. Cruz
Maintained by ARL Web Administrator
Last Modified: September 8, 1997

Scholarly Communication and Technology

Conference Organized by The Andrew W. Mellon Foundation

at Emory University
April 24-25, 1997

Copyright © of the papers on this site are held by the individual authors or The Andrew W. Mellon Foundation. Permission is granted to reproduce and distribute copies of these works for nonprofit educational or library purposes, provided that the author, source, and copyright notice are included on each copy. For commercial use, please contact Richard Ekman at the The Andrew W. Mellon Foundation.

Session #4 Patterns of Usage

Online Books at Columbia: Measurement and Early Results on Use, Satisfaction, and Effect

Carol A. Mandel
Deputy University Librarian
Columbia University

and

Mary C. Summerfield
Coordinator, Online Books Project
Columbia University Libraries

and

Paul Kantor
Consultant

……

4. USE OF ONLINE BOOKS

4.1 Methodology For Studying Use of And Reactions to Various Formats

We laid out the evaluation methodology for this Project in our *Analytical Principles and Design*. This methodology, formulated in the first year of the Project, remains the working plan.

4.1.1 Measurement Plans

Analytical Principles and Design sets forth our plans in this area as follows:

Success of online books is in large part measured by the rate of adoption by the scholarly community and the extent to which they appear to be replacing print books in use. Data on the use of online books and circulation of print books are also available which will allow us to draw certain conclusions on how the various formats are being used.

A related component of our plan is to study the socio-technical environment in which the Columbia community functions and adoption of other forms of electronic communication and scholarly research under the hypothesis that the more Columbia scholars are familiar and comfortable with computing and electronic resources the more likely they are to adopt online books. We summarize some of the early data on this socio-technical environment below. (Section 7 discusses this analysis further.)

4.1.2 Documentation Measures for Use of Online Books

Some of the key measures for documenting use of the online books are:

- The records of the Columbia computing system provide, for the most part, the use data for the online books. For books accessed via the World Wide Web, information on date, time and duration of session involving an online book, user's cohort, location of computer, number of requests and amount of the book requested, means of accessing the book, and networked printing activity will be available. These data will become available in summer 1997 with the full implementation of the authentication system and related databases.

- Circulation data for each print book in the regular collection provides information on number of times a book circulates, circulation by cohort, duration of circulation, number of holds and recalls. For most libraries, the data available for reserve books is the same as that for books in the regular collection as the CLIO circulation system is used for both.

- The records of the Columbia computing system provide, for the most part, the use data for the books accessed via CNet, Columbia's original, gopher-based Campus Wide Information System, including the number of sessions and hits, their date and time. These records do not include the duration of the session, the activity during the session, e.g., printing or saving, or anything about the user. Thus, all we can analyze are the patterns of use by time of day, day of week, and over time.

- Until March 15, 1997, for books accessed via CWeb, we knew the use immediately preceding the hit on the book, the day and time of the hit. For data collected through that point, our analysis is constrained to patterns of use by time of day, day of the week, and over time. By manual examinations of server data, we counted how many hits a user made on our collection during one session and the nature of those hits.

- Since March 15, 1997, we are able to link user information to usage information and derive a series of analyses involving titles used, number of hits, number of books used, and the like by individual and to group those individuals by department, position, and age. These data do not yet include number of sessions of use, just the magnitude of overall use during the period. Session specific data will be available by fall 1997.

4.1.3 Documentation Measures for Reactions to Online Books

We are using a wide range of tools in trying to understand the factors that influence use of online books.

Table 1 summarizes our complex array of surveys and interviews.

Table 1. Types of Surveys

Population	Method	Contact	Rate	Remarks
Users of Online Books	Online instrument	Passive	Low	
Users of Online Books	Online post-use survey	Passive	Very Low	
Users of paper alternatives	Response slips in books	Passive	Unknown	Levels of use not known
Users of course materials in either form	Interviews distributed in class	Active	High	
Users and non-users	Library & Campus-Wide surveys	Active	Moderate	No full active survey of the campus has been done
Discipline-specific potential users	Surveys & Interviews	Active	High	Thus far only conducted before books were online

Note: Passive instruments are ones which the user must elect to encounter. Active instruments are distributed in some way, to the attention of the user. High response rates are in the range of 80-90 percent completion, with better than 60 percent usable.

4.2 Socio-Technical Environment

In our analytical construct, we posit that three sets of socio-technical environmental factors and their change over time will influence the adoption of online books by the Columbia community. These are external (U.S.), disciplinary, and Columbia-related factors. The first and the third of these are discussed below.

4.2.1 External Socio-Technical Environment

In tracking the external socio-technical environment that might affect adoption of online books by members of the Columbia community, we look at three primary measures:

1. Attention to the Internet and related issues in the press, measured by *New York Times* articles;
2. Trends for prices and technical specifications for personal computers, measured both by looking at recommendations for minimum computer standards offered by various writers and at the offerings of Gateway 2000; and
3. Penetration of computers, modems, Internet access into American homes as reported by various market research companies.

Our findings to date are summarized below.

4.2.1.1 Media Coverage Of the Internet

We hypothesize that members of the Columbia community are more likely to feel that up-to-date personal computer systems and online resources are important to their lives and scholarly work the more the media that they see report on them. *The New York Times* is our media proxy in tracking the number of stories that community members might have seen involving online-related topics over the past three years.

Table 2. *New York Times* Stories Involving Information Services

Descriptor Term	**1994**	**1995**	**1996**	**1994 - 1996**	**Pct Chg. '94-'95**	**Pct Chg. '95-'96**
Internet	66	315	360	741	377%	14%
Online Information Services	0	161	140	301	NA	-13%
World Wide Web	0	112	106	218	NA	-5%
Information Superhighway	27	12	5	44	-56%	-58%
Electronic Publishing	30	29	24	83	-3%	-17%
Computer Networks	187	129	46	362	-31%	-64%

Source: *Periodical Abstracts*, using so=New York Times, de=Descriptor Term here, and period=Year given here.

Discussions of the *Internet* soared from 1994 to 1995 and then stayed at a relatively even level of about one story a day. *Online Information Services* and *World Wide Web* went from not even being descriptor terms in *Periodical Abstracts* for 1994 to coverage at about half the rate of the Internet in general in the next two years. These terms seem to have supplanted *Computer Networks* which was a significant term in 1994.

4.2.1.2 Personal Computer Specifications & Pricing Trends

Since the development of personal computers we have seen a continual growth in the quality of the systems on offer and a flat or declining price for the systems recommended for household purchase. In 1997 for the first time, manufacturers have introduced systems priced at around $1,000 that will allow a household to access the Internet smoothly if not with the speed and monitor performance of a system costing twice that much. In June 1997, Gateway 2000 was offering a family-oriented system for $1,499 that was significantly more powerful in almost every parameter than a system priced at $1,999 in May 1996.

Appendix 3 tracks the minimum recommended specifications for home computers given by various writers from May 1994 to April 1997. Summarizing these data by looking at three major factors (CPU, RAM, and hard drive capacity)), we see dramatic increases over the past three years. In the earlier years, neither Pentium CPUs nor personal computer hard drives with capacity above 340 MB were even available.

Table 3. Minimum Recommended Specifications for Home Computers, May 1994 - April 1997

	May 1994 (for student)	April 1995	April 1996	April 1997
CPU	486	486DX2/66	75 Mhz Pentium	166 Mhz MMX Pentium
RAM	4 MB	8 MB	8 MB	16 MB
Hard Drive	100 MB	340 MB	1 GB	2 GB
Price Est.	$1,500	$1,800 - $2,000	$2,000	Not given in the source
Note: This is an extract from Appendix 3.				

As one might expect given Gateway 2000's leading position in the family personal computer market, its offerings track these recommendations by journalists. As Appendix 4 shows, the personal computer capability available for about $2,000 has escalated since late 1994, our first data point. All of these computers are equipped with CD-ROMs, sound systems, and modems. Summarizing that appendix, we find that a $1,500 computer today is over twice as large and twice as fast as a $2,100 computer thirty months ago.

Table 4. Characteristics of a $2,000 Computer, December 1994 - June 1997

	Dec. 1994	**April 1995**	**May 1996**	**May 1997**	**June 1997**
CPU	60 Mhz Pentium	60 Mhz Pentium	120 Mhz Pentium	200 Mhz MMX Pentium	166 Mhz Pentium
RAM	8 MB	8 MB	16 MB	16 MB	16 MB
Hard Drive	540 MB	540 MB	850 GB	1.6 GB	1.2 GB
Price (+ shipping)	$2,099	$2,099	$1,999	$2,064	$1,499
Note: This is an extract from Appendix 4.					

4.2.1.3 Household Computer Penetration & Internet Access

Many market research reports estimate the penetration of computers and modems into U.S. households, access to and use of the Internet, and the like over the past few years. Unfortunately, the findings vary considerably for single points in time (see Appendix 5). Data from one source, Find/SVP, are summarized here.

Find/SVP's Emerging Technologies Research Group issued the results of its latest survey in early May 1997. The telephone survey, conducted from February to April 1997, included 1,000 adult current Internet users and 1,000 adult non-users. Its Web site (http://www.etrg.findsvp.com/internet/) has a substantive summary of its results. The report also summarizes historical penetration data back to 1994 and makes projections through 2001 in a chart (at http://www.columbia.edu/cu/libraries/digital/texts/forecast/) that tracks PC Households, Modem Households, Internet Households, and Non-PC Internet Access Households (NetTV). According to that chart,

- PC households are increasing at a relatively moderate rate - from about 30 million in 1994 to about 37 million in 1997 projected to about 40 million in 1999 and 46 million in 2001. U.S. households number just under 100 million, so these values approximate the household penetration as well - moving from the low 30's to about 46 percent.

- Modem households started out in 1994 as about 40 percent of PC households, but the two values are converging over time to 75 - 80 percent modem penetration in 1997 and a projection of about 95 percent penetration in 1999 and thereafter. Virtually any new household computer purchased from 1997 on will be equipped with a modem.

- Elsewhere, Find/SVP projects a rapid reduction in the market share of modems of less than 28.8 kbps - from 66 percent at year end 1996 to 30 percent at year end 1997 to only 10 percent at year end 1998. They project that 56.6 kbps modems will have 10 percent market share at year end 1997 and 25 percent at year end 1998. These values reflect the sales of modems, not the stock of household computers, which will lag this changeover considerably. This suggests that scholars reliant on modems to access online resources are likely to have relatively slow connections for the next few years. On the other hand, modems are not costly, so if a scholar finds the online resources valuable, he may upgrade to a faster modem.

- Internet households are following a similar pattern of increasing penetration within the universe of modem households. Find/SVP estimates U.S. Internet households as:

Table 5. U.S. Internet Households

Year	Millions of HH	Penetration of Modem HH
1994	3.1	25%
1995	6.2	
1996	14.7	
1997	21.9	75%
1998	28.0	
1999	33.0	87%
2000	36.5	
2001	40.0	93%

- While there are hardly any Non-PC Internet Access Households, i.e., those using NetTV-type systems, now, Find/SVP estimates that there will be about seven million in 1999 and 24 million in 2001.
- Based on telephone surveys, Find/SVP estimates that 8.4 million U.S. adults were current users of the Internet in 1995, 28.8 million in 1996, and 31.1 million in early 1997. They project that 36.3 million adults will be users by year end 1997. Find/SVP asserts that *55 million Americans are poised to become Internet users*. Scholars have a greater exposure to the potential of use of the Internet than do adults in general, so their rate of adoption is likely to be more rapid.
- While Find/SVP found general enthusiasm about the Internet and the Web - about half the current adult Web users use it daily, they also found that nine million Americans have tried the Internet but are not current users.

An early 1997 Baruch College-Harris Poll survey of 1,000 households found 21 percent of U.S. adults (40 million) using the Internet and/or the World Wide Web. This figure is half of all computer users and double the number using the Internet a year ago. An additional 12 percent of respondents use commercial online services.

4.2.2 Columbia Socio-Technical Environment

Columbia infrastructure, penetration of ready access to computing, and amount of time spent in online activities are among the Columbia socio-technical environmental factors that may affect adoption of online books.

4.2.2.1 Campus Infrastructure: February 1997

Columbia's campus infrastructure is similar to that of other universities in its components and in its constant expansion to meet community demand for access to email and other Internet services. Currently, a 10BaseT fiber optic campus network connects 65 buildings and a T3 line connects the campus to the Internet. Over 9,000 ports are connected to the network and over 20,000 computers are registered to community members. All fifteen undergraduate residence halls are pre-wired; the residence hall network has over 4,500 ports. Our modem pool is constantly growing to serve demand; 298 modems with SLIP/PPP support now handle over 52,000 calls on a typical week. Email servers managed over 442,000 email messages in 1996. The campus has 366 public workstations, kiosks, and lab computers; all are connected to the network.

4.2.2.2 Community Perceptions of Access To Computing Resources

Is there a computer (in the library or elsewhere) attached to the campus network (directly or by modem) that you can use whenever you want? is one of two constant questions on our various questionnaires. The most recent response to that question to date came in the Libraries' onsite user survey in March 1997.

- Almost 81 percent of the 2,367 respondents to this question answered *Yes*. This response indicates that, whether they possessed their own computers or not, most community members perceived that they had adequate access to networked resources.
- Looking at the responses by cohorts using individual libraries, we find that the shares responding *Yes* varied from highs of 100 percent for 54 users of the Geoscience Library and 96 percent of 70 users of the Physics Library to lows of 59 percent for 312 users of the Business and Economics Library and 54 percent for 22 users of the Rare Books and Manuscripts Library.
- As the following table shows, there is a statistically significant range to the responses by Columbia status. The particularly small sample of faculty members makes this value suspect. These values vary insignificantly from the equivalent survey a year earlier when the faculty count was 63.

Table 6. March 1997 In-Library Survey: ***Is there a computer (in the library or elsewhere) attached to the campus network (directly or by modem) that you can use whenever you want?***

Cohort	Sample Size	Responding *YES*
Faculty Member	44	86%
Doctoral Student	468	85%
Masters Student	611	67%
Undergraduate	1,065	87%

In Fall 1995, we cooperated with the Office of the Provost in conducting a campus computing survey. The initial means of distributing this survey was an "opinion festival" in the rotunda of the main administration building. This festival was billed primarily as a food tasting; it attracted many students and few faculty members. The computing survey garnered 414 student responses - 125 graduate students and 289 undergraduate students spread fairly well across the four classes. To amplify the graduate student and faculty counts we did follow-up mailings - to a sample of 2,000 graduate students and all faculty members. Responses were modest in number and quite skewed by department, especially for the faculty survey, so these data are unlikely to be reliable.

The share of Columbia community members reporting ready access to a networked-linked computer (the same question asked in the onsite library survey) by cohort is as follows.

Table 7. Fall 1995 Campus Survey: ***Is there a computer (in the library or elsewhere) attached to the campus network (directly or by modem) that you can use whenever you want?***

Cohort	**Sample Size**	**Responding *YES***
Faculty Member	143	90%
Graduate Student	301	80%
Senior	88	65%
Junior	71	63%
Sophomore	76	63%
Freshman	54	78%

With such small sample sizes for the undergraduate cohorts, there is no significant relationship between the shares reporting such computer access and level of study.

About 72 percent of undergraduates, 80 percent of graduate students, and 85 percent of faculty members responded *Yes* to the question *Do you have your own computer in your residence?* in this survey. That these values are higher than those for the access question may reflect that some of the students do not have modems or network cards in their computers or do not use them. Questions asking for details about the power of these computers and the degree to which they have communications hardware were not answered fully.

4.2.2.3 Community Use of Online Resources

A related question that we ask on all of our questionnaires regards time spent on online activities. For the 1996 and 1997 onsite library surveys, this was phrased as *On average this semester, how many hours per week do you spend in online activities (Email, Listservs & Newsgroups, CLIO Plus, Text, Image or Numeric Data Sources, Other WWWeb Uses)?* The respondent was instructed to write a value in the blank provided.

The following table gives a grouping of the distribution of the total responses to this question in 1997 in column 2, of the responses by those who claimed easy access to computers with online access in column 3, and of the responses by those who said that they did not have such access in column 4.

Table 8. March 1997 In-Library Survey: Weekly Hours on Online Activities by Access to Computers Linked to Campus Network, Winter 1997

Hours/Week	Percent of Respondents In Group		
	All (N=2,493)	W/Easy Access (N=1,853)	W/O Easy Access (N=428)
O	2.9%	1.4%	5.6%
1-3	46.4%	45.2%	49.8%
4-6	23.0%	23.8%	21.7%
7-9	6.1%	6.1%	7.5%
10-12	11.7%	12.4%	9.3%
13-15	3.2%	3.8%	1.2%
16-18	0.4%	0.4%	0.2%
19-21	3.7%	4.0%	2.8%
22-28	0.6%	0.4%	1.2%
29-35	1.2%	1.6%	0.2%
More than 35	0.8%	0.9%	0.5%

Even those who answered *No* to the previous question, i.e., they do not feel that they can use a computer attached to the campus network whenever they want, report spending substantial time on online activities each week (column 4 data). The mean number of weekly hours in online activities reported by those who reported any such use was 5.8 hours, with the greatest amount reported 60 hours (8 respondents).

Another way to look at these data is to group the responses by Columbia status of the respondent. This is done below for the four major scholarly components of the community. The cohorts include only those individuals who provided status information. Time spent in online activities was quite consistent across cohorts within the Columbia community; differences among cohorts were not statistically significant.

Table 9.March 1997 In-Library Survey: Weekly Hours In Online Activities by Columbia Status, Winter 1997

	Percent of Respondents In Group			
Hours/Week	Undergraduate Students (N=1,107)	Masters Students (N=649)	Doctoral Students (N=477)	Faculty Members (N=45)
O	2.1%	2.5%	1.7%	6.7%
1-3	49.4%	44.8%	44.2%	33.3%
4-6	22.4%	23.6%	23.1%	26.7%
7-9	6.7%	6.2%	6.3%	0.0%
10-12	10.2%	12.9%	13.8%	17.8%
13-15	2.6%	3.4%	3.8%	8.9%
16-18	0.4%	0.5%	0.6%	0.0%
19-21	3.3%	3.9%	3.6%	4.4%
22-28	0.4%	1.1%	0.6%	0.0%
29-35	1.3%	0.9%	1.7%	1.7%
More than 35	1.3%	0.3%	0.6%	0.0%
Mean	5.7	5.9	6.3	6.5

Differences in reporting make comparison with the 1996 results difficult, but it appears that average weekly hours online increased modestly from winter 1996 to winter 1997.

4.3 Findings On Use Of Books In Online Collection

At this point we will report on (1) trends in use of the CNet and CWeb books; (2); user location and cohort as suggested by host computer address; (3) distribution of use by day of week and time of day; (4) patterns of hits per Web session involving online books for two weeks' use and for the overall use of three social work titles; and (5) use of the online books by individuals from March 15 to May 31, 1997. Summarized below are findings in these areas for the various groups of books.

4.3.1 Reference Books

4:3.1.1 Total Use Over Time

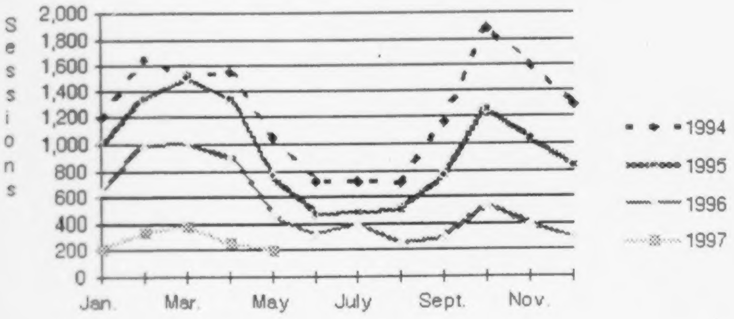
Three reference works have been available online long enough to have generated substantial usage data. These are *The Concise Columbia Electronic Encyclopedia, Columbia Granger's World of Poetry,* and *The Oxford English Dictionary.* The three Garland titles have been online only since the turn of the year or later, so our usage data are very short term for these titles. All three are accessible both through CNet and CWeb.

As of the time of this writing, CWeb usage data extended only through March 14, 1997 on a monthly basis. With the exception of *Columbia Granger's World of Poetry,* usage (number of hits and unique users) from March 15 to May 31, 1997 was reported as a single number. No data are available for *Granger's* after March 14th. In the CWeb data reported below, the early March data is included with the newer data to give one value for the three month period of March to May.

4.3.1.1.1 Concise Columbia Electronic Encyclopedia

The *Concise Encyclopedia* remains on the older CWIS-gopher platform CNet. Usage declined 84 percent over the past three years, from 1,551 sessions in April 1994 to 250 sessions in April 1997. Usage has declined most in the current academic year; 7,861 sessions were registered from September 1995 to May 1996 and 2,941 sessions (63% fewer) from September 1996 to May 1997.

Graph 1. Concise Columbia Electronic Encyclopedia Sessions, 1994 - 1997: CNet



Potential reasons for this steep decline include:

- As community members have become more familiar with the Web, they may be searching it for answers that they might have sought in the *Concise Encyclopedia* when it was our only online encyclopedia.
- Columbia scholars should still be familiar with CNet and the library component, CLIO-Plus, since the library online catalog (CLIO) resides there, but the presence of periodical indexes and the like on the Web has shifted attention away from CLIO-Plus.
- Often encyclopedias on CD-ROM come bundled with new computers; many scholars may own or otherwise have access to these alternatives to the *CCEE*. Those who subscribe to America Online, Prodigy, or Compuserve can use the CCEE or similar resources on those online services.
- Columbia now provides CWeb access to the *Encyclopedia Britannica* (directly from the publisher's server); scholars may be using this instead of the *Concise Encyclopedia*. In December 1996, the Columbia community registered 15,436 hits on the *Encyclopedia Britannica*, up from 8,236 hits in September and 1,096 hits in July.

Columbia scholars seldom use the print copy of the *Concise Encyclopedia*, which resides behind the Reference desk. Its larger cousin, which is out in the public area, sees much greater use. We plan to put that longer, one volume CUP encyclopedia online on CWeb this year. Its use patterns will be instructive.

4.3.1.1.2 Columbia Granger's World of Poetry*

Columbia Granger's World of Poetry is available on both CNet and CWeb. The CNet version is a lynx, non-graphical Web, formulation of the CWeb version. This resource, which became available to the community in online form in October 1994, locates a poem in an anthology by author, subject, title, first line, or keywords in its title or first line. In addition, it provides easy access to the 10,000 most often anthologized poems. As the following table shows, total usage declined from 1996 to 1997 - by 49 percent from the first quarter of 1996 to the first quarter of 1997. Even so, the 4,289 hits for 1996 is considerable.

Reference librarians report no more than a handful of uses of the print version of *Granger's* each year; it is kept behind the main reference desk and lacks the database of poems. The CD-ROM version, which is kept in the Electronic Texts Service, has the same functionality as the online version; it is used once or twice a month on average.

Table 10. Columbia Granger's World of Poetry: Number of Hits by Month

	CNet				CWeb			Total CNet & CWeb			% Change for Total		
	1994	**1995**	**1996**	**1997**	**1995**	**1996**	**1997**	**1995**	**1996**	**1997**	**'94 to '95**	**'95 to '96**	**'96 to '97**
Jan.	0	222	91	18	0	466	150	222	557	168		151%	-70%
Feb.	0	204	137	31	0	282	312	204	419	343		105%	-18%
Mar.	0	292	96	41	0	465	236	292	561	277		92%	-51%
April	0	199	73	34	0	278	NA	199	351	NA		76%	NA
May	0	134	35	17	682	199	NA	816	277	NA		-66%	NA
June	0	81	30		695	102	NA	776	239	NA		-69%	NA
July*	0	80	71		550	383		630	464			-26%	
Aug.	0	78	53		767	27		845	83			-90%	
Sept.	0	76	58		596	179		672	238			-65%	
Oct.	NA	162	84		863	262		1,025	348			-66%	
Nov.	311	114	50		800	413		914	465		194%	-49%	
Dec.	207	68	28		725	257		793	287		283%	-64%	
Total	**NA**	**1,710**	**806**		**5,678**	**3,483**		**6,758**	**4,289**		**NC**	**-37%**	

Note: * July 1995 CNet hits are estimated. CWeb data are available through March 15, 1997 only; this estimated value is twice the actual count.

4.3.1.1.3 The Oxford English Dictionary

At this time, *The Oxford English Dictionary* is the most heavily used reference work in our collection. As noted earlier, it is available on both CNet and CWeb, with the former format having greater functionality but being quite opaque. Users find the latter attractive and easy to use, but it only permits them to look up a definition or browse through the contents.

Usage of the CNet version dropped 59 percent from the fourth quarter of 1994 (2,856 hits) to the first quarter of 1997 (1,167 hits). The CWeb version attracted greater use than the CNet version from its first months. Total usage of the resource was greater with the two versions in place than with only CNet, by 55 percent in February 1997 versus February 1995.

Table 11. Oxford English Dictionary: Number of Hits by Month

	CNet				**CWeb**		**Total CNet & CWeb**		**% Change**		
	1994	**1995**	**1996**	**1997**	**1996**	**1997**	**1996**	**1997**	**94 to '95**	**95 to '96**	**96 to '97**
Jan.	0	643	497	259	0	385	497	644		-23%	30%
Feb.	0	939	1,065	434	0	1,022	1,065	1,456		13%	37%
Mar. *	0	847	683	474	0		683			-19%	
April	0	791	752	372	0	#919	752	2,065#		-5%	#12%
May	0	436	410	300	0		410			-6%	
June	0	336	310		0		310			-8%	
July *	0	300	328		0		328			9%	
Aug.	NA	299	282		8		282			-6%	
Sept.	NA	533	391		570		961			80%	
Oct.	1,238	1,017	783		647		1,430		-18%	41%	
Nov.	975	795	335		271		606		-18%	-24%	
Dec.	643	536	318		337		655		-17%	22%	
Total		**6,926**	**6,154**	**NA**	**NA**	**NA**	**8,069**	**NA**		**-11%**	

Note: * July 1995 CNet usage is estimated, as the true value was unavailable.

# March - May 1997 hits; these data are somewhat under-counted as *The OED* was not included in the user-identified data set initially and as one form of bookmarked access was not included for the whole period. *The OED* became available on CNet in August 1994, but usage data are available back to October 1994 only.

Columbia College has a one semester Logic and Rhetoric course that is required of all its students (about 1,000 each year). Students in this course must complete an assignment involving the *OED* and are encouraged to use an online version. That assignment occurred in October 1996 and mid-February to early March 1997. In the period preceding mid-March 1997,

- almost 42 percent of the hits (1,531) on the CWeb *OED* came from computers in dormitory rooms, suggesting that students are using this resource. This conclusion is confirmed by the analysis of the data by user in the period beginning in mid-March; see section 4.3.4.

Observation and reshelving activity show that scholars frequently use the print copy. However, statistics on use are unavailable as scholars have direct access to several sets in libraries around campus and have not been cooperative in recording use of volumes. In addition, scholars often owned their own copies of the compact edition of *The OED*. Finally, some serious scholars use the CD-ROM version in the Libraries' Electronic Text Service which allows refined searches with a search engine that is more attractive and user friendly than that in CNet.

4.3.1.1.4 Garland Reference Works

Garland's *Chaucer Name Dictionary* was added to the CWeb collection at the end of 1996. *Native American Women* was added in January 1997 and *African American Women* in February 1997. The first two were added to the CNet collection in February 1997 and the third in March 1997.

Table 12. Garland Reference Works: Number of Hits by Month, December 1, 1996 - May 31, 1997

	Chaucer Name Dictionary			*African American Women*			*Native American Women*		
	CWeb	**CNet**	**Total**	**CWeb**	**CNet**	**Total**	**CWeb**	**CNet**	**Total**
Dec. '96	28	NA	28	NA	NA	NC	NA	NA	NC
Jan. '97	62	NA	62	8	NA	8	60	NA	60
Feb.	107	15	122	26	NA	26	107	11	118
March	ND	8	NC		31			7	
April	#72	7	#90	#90	10	#139	#63	4	#77
May		3			8			3	
Total	269	33	302	124	49	173	230	25	255

Note: # March - May 1997 hits. NA - Resource was not available. ND: Data are not available.

NC - Not Calculable.

CWeb is a far more popular means of access to these resources than CNet. Although *Chaucer Name Dictionary* and *African American Women* were both available on CNet from February 3rd, their usage on CNet in February was only 10 to 15 percent of that on CWeb. The Libraries' print copies of these reference books are lightly used, so these hits signify substantial expansion of use of these books.

4.3.1.2 Host Computers for Reference Book Use

A user location analysis acts as a proxy for user cohort for the early use data. We have grouped host computers into the following ten categories.

cc - mainly computers in public labs

cul - computers in the libraries

cunix - in general on campus computers linked directly to a cunix server, also now the host computer for *Granger's*

cupress - computers at CUP

dialup - computers connected by dialup modem

english - computers in the English department

pols - computers in the Political Science department

rhno - computers on the residence hall network

sipa - computers at the School of International and Public Affairs

ssw - computers in offices and labs at the School of Social Work

other - computers at all other Columbia locations

The distribution of use of the five reference works supplied via CWeb across these categories is shown below. With the exception of the three Garland books, a very small share of the uses of these reference works occur on computers in the libraries; the Columbia community is taking advantage of the out-of-library access to these resources. As noted earlier, a large share of the use of *The OED* occurs from students' on campus residences (rhno host computers).

Table 13. Host Computers for Reference Book Use, May 1, 1996 - March 15, 1997 - Percent Distribution

Host Computer Type	*Granger's Poetry*	*OED*	*Garland Titles*
cc	1%	8%	2%
cul	1%	2%	40%
cunix*	64%	16%	**36%**
cupress	**	**	6%
dialup	6%	13%	0%
english	0%	0%	0%
pols	0%	0%	0%
rhno	9%	**42%**	4%
sipa	1%	1%	0%
ssw	7%	1%	3%
other	11%	16%	8%

Notes: * In the later part of this period, a Cunix server was given as the host computer for all uses of *Granger's*.

** Less than .5%

……

Cover Page Table of Contents Next Page Footnotes

Previous Page

For additional information about the conference, or The Andrew W. Mellon Foundation's scholarly communication initiatives, please contact Richard Ekman. For additional information about ARL or this web site contact Patricia Brennan, ARL Program Officer at (202) 296-2296.

Return to Office of Scholarly Communication Home Page

ARL Home ARL Scholarly Communication and Technology Home Page

© Association of Research Libraries, Washington, DC
Web Design by Angelo F. Cruz
Maintained by ARL Web Administrator
Last Modified: September 8, 1997

Scholarly Communication and Technology

Conference Organized by The Andrew W. Mellon Foundation

at Emory University
April 24-25, 1997

Copyright © of the papers on this site are held by the individual authors or The Andrew W. Mellon Foundation. Permission is granted to reproduce and distribute copies of these works for nonprofit educational or library purposes, provided that the author, source, and copyright notice are included on each copy. For commercial use, please contact Richard Ekman at the The Andrew W. Mellon Foundation.

Session #4 Patterns of Usage

Online Books at Columbia: Measurement and Early Results on Use, Satisfaction, and Effect

Carol A. Mandel
Deputy University Librarian
Columbia University

and

Mary C. Summerfield
Coordinator, Online Books Project
Columbia University Libraries

and

Paul Kantor
Consultant

. . . .

4.3.2 Monographic Books

4.3.2.1 Total Use Over Time

The Online Books Project includes three collections of monographic books for which we now have some use data. These are (1) *Past Masters,* classical texts in social thought; (2) *Columbia University Press Monographs,* mostly contemporary social work books; and (3) *Oxford University Press Monographs,* contemporary philosophy and literary criticism books. Most of these books came online during the 1996-97 academic year.

4.3.2.1.1 The Past Masters Collection

Until July 1996, ten *Past Masters* texts were available to the Columbia community online; since then, 54 texts have been available.

As Table 14a shows, from September 1996 to May 1997, the *Past Masters* texts registered about 2,460 hits from the scholarly community. Table 14b displays the number of hits on the eight most heavily used of these texts for the period from September 1996 to May 1997. This group of texts registered 1,692 hits from the Columbia community, or about 69 percent of the total usage for the *Past Masters* for this period. Thus, in a collection of texts that was not specifically selected to meet the specific needs of a set of users, we find that 15 percent of the texts accounted for 69 percent of the usage. The other 46 texts averaged about 17 hits each over this period, or about two hits per month.

Patterns of usage may be expected to change over time as various texts are used in courses or by researchers and as the Columbia community becomes more aware of the online books. It will be interesting to see how usage of the *Past Masters* evolves over the next academic year. The data to date remind us that to the extent that there are meaningful costs to creating online books (or journals) and to maintaining them as part of a library's collection, planners must select items for the online collection carefully. Of course, the decision rules for a consortial approach will be different from those for a group of non-cooperating individual libraries. We are attempting to delve into these cost issues and hope to have some findings by the end of 1997.

Table 14a. *Past Masters* On The Web, Total Monthly Hits: May 1995 - March 1997

Month	1995	1996	1997
January	NA	231	236
February	NA	298	334
March	NA	137	
April	NA	1,423	#205
May	0	144	
June	5	1,861	
July	0	513	
August	52	340	
September	0	330	
October	972	369	
November	619	604	
December	225	382	

Note: Early data include a substantial share of Project development hits as work was progressing on the texts. # March - May 1997 hits.

Table 14b. Key *Past Masters* Texts On The Web, Monthly Hits: August 1996 - May 1997

Author	Bacon	Bentham	Hamilto	Hobbes	Hume	Machiavel	Mill		
Title	***New Organon***	***Introduction Principles of Morals & Legislation***	***Federalis Papers***	***Leviatha***	***Enquiry - Principles of Morals***	***The Prince***	***On Liberty***	***US Constitutio***	**TOTAL**
Sept. '96	45	1	23	21	5	61	1	35	192
Oct.	27	18	18	54	12	75	18	15	237
Nov.	321	4	1	**74**	2	**76**	5	16	499
Dec.	14	36	5	**99**	18	68	8	13	261
Jan. '97	63*	1	11*	9	60	30	2*	4	180
Feb.	34*	20	**13**	16	**39**	40*	**34**	16	212
March	*	0	*	1	12	*	*	5	18
April	#37*	0	#14*	0	0	#20*	#22*	0	#93
May	*	0	*	0	0	*	*	0	NC
Total	541	80	13	274	148	310	90	104	1,692

Note: Values that are in **bold** indicate that the text was in use in the Columbia College Contemporary Civilization class in that month. Other Past Masters texts used in that course (Locke's *Second Treatise on Civil Government)* did not make our heavy use list. * This text was on reserve for one or more courses during this semester. # March to May 1997 hits.

4.3.2.1.2 Columbia University Press Monographs

Cross title comparisons are difficult because books were made available to the community at different times - from September 1996 forward. Our design breaks these books into chapter files in most cases, so a hit gives a user access to a whole chapter if he gets beyond the Table of Contents file. As Table 15 shows, scholars are using these online books.

4.3.2.1.2.1 Social Work Books

In the period from May 1, 1996 to May 31, 1997, the social work books received a total of 1,948 hits, with a peak in October of 353 hits. The October peak reflects the use of the three books with the most hits in classes in the School of Social Work. **Bold** values on Table 15 indicate months in which we are aware that the book was being used in a class. If faculty members did not put the books on reserve in the library, we may not know that it was in use in a class. Also in many cases, although we know the book was used in a course, we do not know in which months. The secondary peak in February 1997 (278 hits on social work books) also reflects class use of the two titles with the greatest number of hits.

We informed the social work faculty of the availability of the online availability of these books in several ways over the months preceding their introduction. Furthermore, we requested permission of these instructors to conduct in-class surveys at the time they were discussing the material from these books. These steps seem to have led the instructors to inform their students of the availability of these books online and to have caused some faculty and students to look at the online books. In the Spring 1997 term, we also provided handouts about the Social Work collection to several classes that were using books included in the collection.

Table 15. Scholarly Hits on Columbia University Press Books, May 1, 1996 - May 31, 1997

	1996								1997					1996-97
Title	**May**	**June**	**July**	**Aug**	**Sept**	**Oct.**	**Nov**	**Dec.**	**Jan.**	**Feb.**	**Mar**	**Apr.**	**May**	**Total**
1. Handbook of Gerontological Services	0	0	0	4	15	**92**	18	12	13	21	#36			211
1. Philosophical Foundations...	9	3	0	7	15	19	5	11	14*	18*	*#25			126
1. Supervision In S.W.	16	11	1	3	23	**58**	47	13	31	**55**	***#42**			300
1. Task Strategies:	0	1	6	5	18	**150**	109	12	8	2	#288			599
1. Turning Promises ...	0	5	0	6	18	19	10	2	3	20	#32			115
3. Mutual Aid Groups, .	0	1	16	8	9	0	8	59	83	**151**	***#149**			335
5. Qualitative Research In S.W.	0	0	16	4	6	1	3	8	54*	11*	***#10**			113
Total Social Work	25	22	39	37	104	339	200	117	206	278	#356			1799
2. Seismosaurus	0	0	0	0	0	26	55	11	12	39	#10			153
5. Sedimentographica	0	0	0	0	0	0	0	0	29	4	#6			39
Total Earth Science	0	0	0	0	0	26	55	11	41	43	#16			192
3. Gender in Inter'l Relations	0	0	0	0	0	0	5	12	10	6	#22			55

Note: Numbers preceding titles indicate the month the book was made openly accessible to the community. 1=Sept. 1996 and on through the months. **Bold values** indicate months in which the book is known to have been used in a course; * indicates that the book was on reserve for one or more courses during the semester, but month(s) of use are not known. # March - May 1997 hits.

The following books had no hits by the end of May 1997: *Ozone Discourses, Jordan's Inter- Arab Relations, Hemmed In, Managing Indonesia.*

4.3.2.1.2.2 Other Works

The two earth and environmental science titles had 192 hits; the political science title received

55 hits. While these values do not seem large, they should be thought of in the context of the number of uses that a print copy would receive in a similar period if it were not on reserve for a course. If it is recalled almost at once, a book will circulate at most about six times during a semester. Books on loan are also unavailable for serendipitous use by scholars. For example, the two earth and environmental science titles had a total of only two circulations over the past three years.

As the following table shows, the paper copies of these books have experienced substantial circulation, some regular and some reserve, over the past three years. It does not seem that circulation of the print copy has declined with the introduction of the online versions. In fact, it is likely that online availability has created an expanded audience for at least some titles. Further analyses of the print circulation data will be conducted during summer 1997 to determine if additional expansion of use or new shifts in use can be discerned.

Table 16. Columbia Circulation of Columbia University Press Monographs: 1994 - 1996

	1/1994	2/1994	1/1995	2/1995	1/1996	21996	Total
Handbook Geront. Services	3	6	0	18	8	7	42
Mutual Aid Groups,..	0	19	17	16	7	9	68
Qualitative Research In Social Work	0	1	5	4	1	1	12
Supervision In Social Work	39	3	47	4	30	4	127
Task Strategies:	0	15	8	0	15	18	56
Philosophical Foundations of S. W.	0	3	2	1	1	1	8
Turning Promises Into Performance	3	3	7	3	6	1	23
*Sedimentographica**	0	0	0	0	0	1	1
Seismosaurus	0	1	0	0	0	0	1
Gender in Internatl Relations	3	8	13	2	7	5	38
Note: * Title acquired in 1995. Periods are half years: January - June and July-December.							

4.3.2.1.3 Oxford University Press Monographs

The first Oxford monographs were introduced to the Columbia community in mid-Fall 1996. Six books of literary criticism and 12 of philosophy were online by June 1997; Tables 17a and 17b detail the month of introduction and the usage for each. The literary criticism titles received 92 hits through May 31, 1997; the philosophy titles 626 hits.

None of these books was on reserve for a course Fall 1996 or Spring 1997. This is not surprising. Few monographs in the Libraries collection are on reserve for courses. Also, faculty members may take a while to become acquainted with newer monographs and to decide to include them in a course. Potentially, a great value of placing new monographs online will be in helping scholars to maintain current awareness in their fields of scholarship and teaching.

Table 17a. Scholarly Uses (Hits) Oxford University Press Monographs In Literary Criticism, May 1, 1996 - May 31, 1997

Author	*May*	*Lowrey*	*Moses*	*Morrison*	*Spiegel-man*	*Parkes*	
Title	*2. Children's Literature & Critical Theory*	*2. "Littery Man"*	*2. Novel & Globalizati of Culture*	*2. Poetics of Fascism*	*4. Majestic Indolence*	*6. Modernism & Theater of Censorship*	*Total*
May '96	0	0	0	0	0	0	0
June	0	0	0	0	0	0	0
July	0	0	0	0	0	0	0
August	0	0	0	0	0	0	0
Sept.	0	0	0	0	0	0	0
Oct.	0	0	0	0	0	0	0
Nov.	0	3	0	2	0	0	5
Dec.	6	0	11	0	0	0	17
Jan. '97	5	7	4	6	12	0	34
Feb.	1	7	8	7	1	0	24
March			0	0		0	
April	#7	#3	0	0	#2	0	#12
May			0	0		0	
Total	19	20	23	15	15	0	92

Note: Numbers in column headings stand for the month the book entered the public online collection: 1. October 1996 2 November 1996 3. December 1996 4. January 1997 5. February 1997 6. June 1997.

Earlier uses are by individuals informed of the URLs. # Hits for March - May 1997.

Table 17b. Scholarly Uses (Hits) Oxford University Press Monographs In Philosophy, May 1, 1996 - May 31, 1997

	1996								1997					1996-97
Title	**May**	**June**	**July**	**Aug**	**Sept**	**Oct.**	**Nov.**	**Dec.**	**Jan.**	**Feb.**	**Mar**	**Apr.**	**May**	**Total**
Autonomous Agents	0	0	0	0	5	0	0	5	2	20	#3			35
Bangs, Crunches, Whispers & Shrieks	0	1	16	8	11	0	4	8	8	7	#26			89
Free Public Reason:	0	0	0	0	8	0	2	7	0	11	#32			60
Freedom & Moral Sentiment	0	0	0	0	5	0	0	2	2	1	0			10
Law & Truth:	0	0	0	0	0	0	9	0	7	5	#4			25
Morality, Normativity & Society	0	0	2	6	9	0	4	1	14	33	#1			70
Nietzsche's System	0	3	0	0	7	0	0	7	7	2	#9			35
Other Minds: Critical Essays	2	3	5	0	5	0	15	9	12	23	#39			113
Philosophy of Mathematics...	0	0	0	0	8	0	2	3	2	0	#6			21
Real Rights	0	0	3	0	8	0	0	0	5	3	#3			22
Self Expressions:	4	24	14	2	7	0	4	2	9	30	#21			117
The Logic Of Reliable Inquiry	0	3	0	0	6	1	2	4	3	1	#9			29
Total	6	34	40	16	79	1	42	48	71	136	#153			626

Note: Numbers preceding titles indicate the month a book was made openly accessible to the community. Earlier use is by individuals who knew the non-public addresses for the books. All of these books, with the exception of *Law & Truth,* became available to the community in October 1996. *Law & Truth* became available in November 1996. # Total hits March - May 1997.

Table 18. Columbia Circulation of Oxford University Press Monographs: 1995 - 1996

	1/1995	2/1995	1/1996	2/1996	Total
Children's Literature & Critical Theory					
*"Littery Man"**	0	0	0	1	1
Majestic Indolence	0	1	0	0	1
Novel & Globalization of Culture	0	1	3	0	4
Poetics of Fascism	0	0	3	0	3
Autonomous Agents	0	2	1	2	5
Bangs, Crunches, Whispers & Shrieks					
Free Public Reason:					
Freedom & Moral Sentiment	0	0	1	1	2
Law & Truth:					
Morality, Normativity & Society	0	1	0	2	3
*Nietzsche's System**	0	0	1	2	3
Other Minds: Critical Essays	0	1	0	0	1
Philosophy of Mathematics...					
Real Rights					
*Self Expressions: Mind, Morals &...**	0	0	0	3	3
The Logic Of Reliable Inquiry					
Note: * Title acquired in 1996. Data for the titles with blank cells were not collected in the report on circulation, either because they had not circulated through December 31, 1996 or because the Library Systems Office did not include this book in the report; they will be provided in the report in July 1997. A review of the online catalog shows that most of these books have circulated recently.					

Although data on the Columbia circulation of the paper copies of these Oxford books are incomplete at this point, it is clear that these books have some interest for the Columbia community even though they are not on reserve for courses. They are circulating, while some other books sit on the shelf for years before someone checks them out.

One title that did not circulate greatly (*Other Minds*) was held by one scholar for 347 days from late 1995 through much of 1996, thus depriving other members of the community of the opportunity to encounter it, to determine whether it might be of value to their work, and to read it closely. As noted earlier, a key advantage of online books is their ready availability to the whole community at all times. The online version of *Other Minds* received 39 hits in 1996.

4.3.2.2 Host Computers for Monographic Book Use

The host computer categories used in analyzing the location of use of the various books were defined earlier. Looking at the Columbia University Press and Oxford University Press books as a whole, we find the distribution given below.

Table 19. Host Computers for Monographic Book Use, May 1, 1996 - May 31, 1997

Host Computer Type	Columbia University Press	Oxford University Press
cc	2%	5%
cul	19%	33%
cunix	11%	15%
cupress	12%	2%
dialup	11%	17%
english	0%	1%
pols	0%	0%
rhno	1%	3%
sipa	1%	*
ssw	41%	12%
other	5%	12%
Note: * Less than 0.5%.		

The detailed data on the host group for each book in the collection confirms what one would expect from these data -- host computer type is related to book type for the most part. However, once a group within the community becomes aware of the online books, they are likely to review other books in the collection (at least in this early stage when the collection is small). For example, half of the use of *Autonomous Agents: From Self Control to Autonomy* was from social work host computers. This is a title that might seem related to social work issues even though it is not one of the Columbia University Press social work books or part of the collection of the Social Work Library.

SSW was the host location for the following shares of the hits on the social work titles:

Handbook of Gerontological Services	53%
Mutual Aid Groups, Vulnerable Populations	76%
Philosophical Foundations of Social Work	39%
Qualitative Research in Social Work	69%
Supervision in Social Work	48%
Task Strategies: An Empirical Approach	68%

Closer analysis of the usage data finds substantial use from the computer lab in the Social Work School as well as from faculty computers. This suggests that many of these graduate students, most of whom do not live on or near campus, may not have Web access in their homes and, hence, at this point in time, are not equipped to take full advantage of the online books from home. Use of the online version enables them to use the books from the School, however, thus avoiding the walk of several blocks to the Social Work Library.

4.3.2.3 Use By Day And Time -- All Types of Books

Table 20 gives the breakdown of use of the online materials by day of the week for May 1, 1996 to March 15, 1997. Table 21 gives the breakdown by time of day for the same materials for the same period.

The patterns of use varied considerably among the families of online books. For example, 79 percent of the use of *The OED*, 80 percent of the use of the Oxford monographs, and 91 percent of the use of Columbia monographs occurred on weekdays. Friday alone accounted for 25 percent of the hits on Oxford monographs and 20 percent of the hits on Columbia monographs. This concentration of use is not surprising, as few classes meet on Friday at Columbia, making it a good day for both faculty and students to do research and class assignments. We will track future data to see if these patterns continue.

Table 20. Patterns of Use from Web Server: May 1, 1996 - March 15, 1997 Hits

	Granger's Poetry		**Past Masters Texts**		**Oxford English Dictionary**		**Garland Books**		**OUP Monographs**		**CUP Monographs**	
Day of Week	**Num ber**	**% of Total**	**Num ber**	**% of Total**	**Num ber**	**% of Total**	**Num ber**	**% of Total**	**Num ber**	**% of Total**	**Num ber**	**% of Total**
Monday	360	15%	1,338	26%	739	20%	137	26%	74	11%	247	12%
Tuesday	361	15%	1,090	21%	597	16%	144	27%	85	13%	439	20%
Wednesday	456	19%	741	14%	619	17%	93	18%	101	15%	416	19%
Thursday	273	11%	711	14%	447	12%	43	8%	101	15%	429	20%
Friday	486	20%	672	13%	484	13%	51	10%	166	25%	423	20%
Saturday	198	8%	224	4%	266	7%	6	1%	55	8%	99	5%
Sunday	268	11%	420	8%	520	14%	56	11%	74	11%	96	4%
Total	2,402	100%	5,196	100%	3,67	100%	530	100%	656	100%	2,149	100%
Weekdays:	1,936	81%	4,552	88%	2,88	79%	468	88%	527	80%	1,954	91%

Table 21. Hits by Time of Day: May 1, 1996 - March 15, 1997

	Granger's Poetry		Past Masters Texts		Oxford English Dictionary		Garland Books		OUP Monograph		CUP Monographs	
Time Period	**Number**	**% of Total**	**Number**	**% of Total**	**Number**	**% of Total**	**Number**	**% of Total**	**Number**	**% of Total**	**Number**	**% of Total**
Midnight - 6: AM	243	10%	375	7%	560	15%	65	12%	69	10%	46	2%
6: AM - Noon	365	15%	1,193	23%	405	11%	125	24%	96	15%	381	18%
Noon - 6: PM	1,275	53%	2,199	42%	1,288	35%	259	49%	386	59%	1,275	59%
6: PM - Midnight	519	22%	1,429	28%	1,419	39%	81	15%	105	16%	447	21%
9: PM - 9: AM	482	20%	1,202	23%	1,271	35%	122	23%	115	18%	148	7%
9: AM - 9: PM	1,920	80%	3,994	77%	2,401	65%	408	77%	541	82%	2,001	93%

The time of day analysis finds:

- The use of reference books occurred mostly in hours in which the libraries are typically open, i.e., between 9: AM and 9: PM, but a meaningful share - 20 percent (*Granger's)* to 35 percent (*OED)* - occurred from 9: PM to 9: AM.

- The use of online monographs occurred almost totally (93 percent for the Oxford monographs and 95 percent for the Columbia monographs) in hours in which the libraries are typically open, i.e., between 9: AM and 9: PM. Users of these books do not seem to have been taking advantage of the constant availability of online materials. This suggests that these books may have been receiving a large share of their use from computers in the libraries or elsewhere on campus, such as the computer lab in the School of Social Work, that are used predominately or exclusively during work hours. The distribution of use by host type is discussed in the preceding section.

The online versions of these books provide scholars with the flexibility of access to materials at times of the day and week when they cannot use them in the libraries, either because the libraries are not open or because the scholars are not able or willing to be in the library at that time. This flexibility is likely to enhance the scholar's efficiency and effectiveness, but use patterns do not yet indicate that it is being exploited.

4.3.3 Session Analysis for Use of CWeb Books

4.3.3.1 Two Weeks' Sessions - All Online Books

We extracted the Web session data for the online books for the weeks of October 26 and December 7, 1996, in order to learn about the number of sessions and the number of text hits per session. The analysis involved looking for what seemed to be sequential hits from the same address, i.e., ones that were very close in time, and counting those as part of a session of using online books.

The number of sessions in the second week (212) was one-third greater than in the first week and the number of hits (611) was twice as great. The mean session included about two hits in the first week and three hits in the second. Several sessions seemed to involve systematic retrieval of many files.

One way to put this usage into perspective is to compare it to use of other library-related electronic services. In December 1996, bibliographic indexes on CLIO-Plus, the Libraries component of CNet, had a total of 16,740 hits (or about 4,000 per week). However, individual indexes sustained monthly usage ranging from 28 hits (AGELINE) to 2,743 hits (MEDLINE). In fact, only MEDLINE sustained an average number of hits per week that was greater than the number of hits to online books in the December 1996 sample week.

Table 22. Online Book Usage (Web): Hits Per Use Session

	Week of Oct. 26, 1996			Week of Dec. 7, 1996		
Hits Per Session	**# of Sessions**	**% of Sessions**	**# of Hits**	**# of Sessions**	**% of Sessions**	**# of Hits**
1	90	58%	90	102	48%	102
2	38	25%	76	56	26%	112
3	9	6%	27	17	8%	51
4	2	1%	8	9	4%	36
5	4	3%	20	4	2%	20
6	4	3%	24	4	2%	24
7	2	1%	14	3	1%	21
8	4	3%	32	6	3%	48
9	0	0%	0	2	1%	18
10-15	1	1%	11	3	0%	39
16-20	0	0%	0	2	1%	36
21-25	0	0%	0	0	0%	0
>25	0	0%	0	4	2%	104
TOTAL	**154**	**100%**	**302**	**212**	**100%**	**611**

4.3.3.2 Sessions for Social Work Books

As noted earlier, in Fall 1996, three social work books were most intensively used as they were assigned reading for courses. We analyzed the server statistics through the end of 1996 for these books in an effort to learn how deeply the books were used - to what extent use sessions included book chapters, the search engine, the pagination feature, and the like.

Looking at each of these three titles, we find:

- Relatively few sessions (7% - 24%) involved someone going to the Table of Contents/Title page for a book and stopping.
- Many sessions (28% - 59%) involved use of more than one chapter of the book; sessions averaged 1.4 to 3.5 hits on chapters, depending on the book used.
- Some users would seem to be repeat users who had bookmarked a chapter in the book or made a note of the URL as some sessions (9% - 17%) did not include a hit on the Table of Contents/Title page.

Summary data follow:

Table 23. Session Analysis for Social Work Books, Fall 1996

	Handbook of Gerontological Services		***Supervision in Social Work***		***Task Strategies***	
Number of Chapters/Essays in Book	10 chapters		10 chapters		11 chapters	
Sessions	41		46		58	
Hits	128		128		284	
Mean Hits/Session	3.1		2.8		4.9	
Sessions with Table of Contents Hit Only	4 (10%)		11 (24%)		4 (7%)	
Sessions with No Table of Contents Hits	7 (17%)		7 (15%)		5 (9%)	
Sessions with >1 Chapter Hits	15 (37%)		13 (28%)		34 (59%)	
TOTAL HITS ON:	**Number**	**Average/ Session**	**Number**	**Average/ Session**	**Number**	**Average/ Session**
Table of Contents	38	.9	47	1.0	60	1.0
Chapters	74	1.8	63	1.4	202	3.5
Page Locator	5	.1	3	.1	7	.1
Search	5	.1	10	.2	4	.1
Bibliographic Page	5	.1	4	.1	9	.2
Author Biography	1	*	1	*	2	*
Note: * Less than .05.						

4.3.4 Analysis of Unique Users Of Online Books

As of March 15, 1997, the online books system required users to sign in with email address and password, for all of our collection except *The OED* and *Granger's Index to Poetry*. This information can be combined with that obtained from the university's directory of its students and staff to obtain demographics on individuals who are using our books. Information on session behavior will be available late summer 1997. Summarized below are findings for the first period of 11 weeks, from March 15 - May 31, 1997, under the new system. Final exams for the Columbia spring semester ended on May 16th with graduation on May 21st. Thus, most of the period under analysis was a busy part of the academic year with only the last few days part of the early summer lull.

4.3.4.1 Use and Users

In this period, the collection (absent *The OED* initially) was used by 280 different persons making 1,439 hits, for an average of over five hits per user. One or more persons used 45 of the books in the collection. Looking at two data periods, we find use breaks down as follows:

Period	**Hits**	**Mean Hits/Week**	**Unique Users**	**Mean Hits/User**
March 15-April 14	591	138	107	5.5
April 15-May 31	848	126	173-280	3.0-4.9
March 15-May 31	1,439		280	5.1

Without a breakdown by individual, we cannot know the overlap between the 107 users in the first month (information obtained in an early analysis) and the 280 users over the whole period. We do know that the mean hits per user decreased from 5.5 in the first month to 5.1 for the whole period.

Comparing the number of hits on each book category for the two periods of data gathering, we find:

Table 24. Hits by Book Category, March 15 - May 31, 1997

	March 15 - April 14			**April 15 - May 31**		
Book Category	**Hits**	**% of All Hits**	**Mean Hits/Wk**	**Hits**	**% of All Hits**	**Mean Hits/Wk**
Garland	24	4%	5.6	83	10%	12.4
Granger's	NA	NA	NA	NA	NA	NA
OED*	70	12%	16.3	417	50%	62.1
Past Masters	94	16%	21.9	31	4%	4.6
Oxford Univ. Press	34	6%	7.9	76	9%	11.3
Columbia Univ. Press	369	62%	86.1	228	27%	34.0
TOTAL	591	100%	137.9	835	100%	124.4

* *The OED* was added late in the first period. One means of access for *The OED* is still not included.

As *The OED* came into the user-based analysis late, it had an artificially low number of hits and share of the total in the first period. As a result, the shares for the other book categories were inflated. In the second period, *The OED* has the prominence among the online books - a 50 percent share of hits - that the other server data have shown, even though one means of access is not included.

During the second period, use of the three Garland reference books doubled to an average of four hits per week per book and hits on the Oxford monographs increased 43 percent. Weekly hits on the *Past Masters* decreased 79 percent and on the Columbia monographs 60 percent. Overall, average weekly hits were down ten percent.

4.3.4.2 Use Concentration

In this section we will analyze the data on the number of users, the amount of use per user, and the demographics of the user population for the various books. This analysis may shed light on the patterns of use and what factors favor use of online books.

4.3.4.2.1 Reference Books

The number of unique users, hits and mean hits per user for each of the reference titles for which data are available during this period was as the following table shows:

Table 25. Unique Users and Hits for Reference Books, March 15 - May 31, 1997

Title	Users	Hits	Mean Hits/User
The OED	173	487	2.8
Chaucer Name Dictionary	9	34	3.8
Native American Women	9	39	4.3
African American Women	6	33	5.5

These data show an inverse correlation between number of users and the mean number of hits per user.

4.3.4.2.2 Non-Reference Books

Half of the 82 online non-reference texts, including the *Past Masters*, were used during this 11 week period. The distribution of titles by number of users was as follows:

Table 26. Non-Reference Books by Number of Users, March 15 - May 31, 1997

Number of Users	Number of Titles	Share of Titles
30	1	1%
18	1	1%
8	2	2%
7	2	2%
6	3	4%
5	0	0%
4	4	5%
3	3	4%
2	10	12%
1	15	18%
0	41	50%

In declining order of number of users, the non-reference texts which had two or more unique users during this period, their number of users, number of hits, and mean hits per user were:

Table 27. Unique Users and Hits for Non-Reference Books, March 15 - May 31, 1997

Title	Users	Hits	Mean Hits/User
Task Strategies: An Empirical Approach to Clinical Social Work	30	288	9.6
Mutual Aid Groups, Vulnerable Populations and the Life Cycle	18	138	7.7
Philosophical Foundations of Social Work	8	21	2.6
Supervision in Social Work	8	33	4.1
Self Expressions: Mind, Morals and the Meaning of Life	7	21	3.0
Bangs, Crunches, Whimpers, & Shrieks	7	21	3.0
Handbook of Gerontological Services	6	31	5.2
Qualitative Research in Social Work	6	10	1.7
Turning Promises into Performance: The Management Challenge	6	31	5.2
Gender in International Relations	4	6	1.5
Other Minds	4	34	8.5
Seismosaurus	4	8	2.0
The Prince	4	12	3.0
Nietzshe's System	3	6	2.0
The Logic of Reliable Inquiry	3	7	2.3
The New Organon	3	24	8.0
Art of Rhetoric	2	8	4.0
An Essay toward a New Theory of Vision	2	2	1.0
Federalist Papers	2	9	4.5
Free Public Reason: Making It Up As We Go..	2	6	3.0
On the Principles of Political Economy...	2	6	3.0
Philosophy of Mathematics & Math. Practice in the 17th Century	2	3	1.5
Principles of the Most Ancient & Modern Philosophy	2	8	4.0
Real Rights	2	3	1.5
Reflections on the Revolution in France	2	5	2.5
Sedimentographica	2	6	3.0

Note: Titles in **bold** were on reserve for one or more courses in Spring 1997.

The two books with the most users, both social work texts, had the most hits and the first and fourth highest mean hits per user. They accounted for 426 (51%) of the 832 hits on non-reference texts during this period. The top four texts (five percent of the non-reference collection), all social work books, accounted for almost 58 percent of the non-reference hits. The mean hits per user are highly variable. Only six texts averaged more than five hits per user.

4.3.4.3 User Cohorts

4.3.4.3.1 Reference Books

The top user departments and Columbia statuses for the reference books are as follows:

4.3.4.3.1.1 The OED

The seven departments that were the source of four percent or more of the users were:

Department	**User Share**	**Hits Share**
Columbia College	21%	18%
Unidentified User (Barnard, Teachers College)	15%	16%
English	9%	13%
Engineering	8%	8%
Computer Science	6%	9%
General Studies	4%	4%
History	4%	2%
Total	67.00%	70.00%

Primary Columbia statuses that were the source of four percent or more of the users were:

Columbia Status	**User Share**	**Hits Share**
Undergraduate Student	58%	55%
Unidentified User	15%	16%
Graduate Student	6%	11%
Professional Student	4%	4%
Total	83%	86%

Faculty were responsible for a total of less than three percent of the hits on *The OED*.

4.3.4.3.1.2 Grangers Reference Works:

4.3.4.3.1.2.1 Chaucer Name Dictionary

The distribution of the nine unique users by department is:

Department	Number of Users	Share of Users
Unidentified User	2	22%
Engineering	2	22%
Columbia College	1	11%
English	1	11%
History	1	11%
Music	1	11%
General Studies	1	11%

The distribution of the nine unique users by primary Columbia status is:

Undergraduate Student	67%
Unidentified User	22%
Associate Professor	11%

4.3.4.3.1.2.2 Native American Women

The distribution of the nine unique users by department is:

Department	Number of Users	Share of Users
Unidentified User	4	44%
Columbia College	1	11%
Genetics	1	11%
Library Systems	1	11%
Social Work	1	11%
"0" *	1	11%
* To be identified		

The distribution of the nine unique users by primary Columbia status is:

Unidentified User	44%
Undergraduate Student	22%
Graduate Student	22%
Catchall Other	11%

4.3.4.3.1.2.3 African American Women

The distribution of the six unique users by department is:

Department	Number of Users	Share of Users
Unidentified User	3	50%
Electrical Engineering	1	17%
Genetics	1	17%
English	1	17%

The distribution of the six unique users by primary Columbia status is:

Graduate Student	33%
Undergraduate Student	17%
Unidentified User	50%

4.3.4.3.2 Non-Reference Books

Almost 91 percent of the users of the top four books, all social work titles, were from the School of Social Work; they accounted for 98 percent of the hits on those books. The vast majority of these users (56 of 64) were graduate students. With the exception of the most used one, *Task Strategies*, these books were on reserve for social work courses during the spring 1997 semester.

- Three sections, with a total of about 70 students, used *Supervision in Social Work* as a key text. Thus, potentially, if all seven graduate student users were members of these course sections, about 10 percent of the students used this book online during this half semester.

- A different three sections, again with about 70 students in total, used *Mutual Aid Groups*. This book was a major reading in the course; in fact, one of its authors taught two of the sections of the course in which it was used. Sixteen graduate students used this title for a potential penetration of about 23 percent.

- *Philosophical Foundations...* (as well as *Qualitative Research in Social Work*) was on reserve for a doctoral seminar which had an enrollment of 11 students. The instructor reported that this book was a major text in the course that students would have bought traditionally. She did not know how many of her students used the online version. If all eight graduate student (7) and professional student (1) users were class members, that suggests a substantial penetration for that small class. However, it is likely that some of these users were not enrolled in that course.

- We have no explanation for the heavy use of *Task Strategies* (by 26 graduate students). The instructor for the course in which the book had been assigned the previous semester reported that she had not recommended it to her students.

The fifth and sixth most used titles - *Self Expressions: Mind, Morals and the Meaning of Life* and *Bangs, Crunches, Whimpers, & Shrieks* - are both Oxford University Press philosophy titles.

- *Self Expressions* is listed in the Current Social Science Web page along with the social work titles. Five of its seven users were from the School of Social Work, one from the Center for Neurobiology and Behavior, and one from Electrical Engineering. Five of the users were graduate students, one an undergraduate student, and one a post doctoral research fellow.
- *Bangs, Crunches, Whimpers, & Shrieks* is listed under *Physics* in the Current Science Web page. Two of its seven users were from the Physics department, another two from unidentified departments, and one each from Electrical Engineering, Engineering and General Studies. Five of the users were undergraduate students and two unidentified status.

Looking at the various non-reference collections overall, we find the following cohort dominance patterns:

4.3.4.3.2.1 Past Masters

The departments with four percent or more of the 125 hits on *Past Masters* were:

Department	Hits	Share of Hits
Economics	29	23%
Anthropology	16	13%
Political Science	16	13%
Visual Arts	13	10%
Unidentified	10	8%
History	8	6%
Columbia College	7	6%
English & Comparative Literature	7	6%
Physics	5	4%
Total	111	89%
Note: Detail may not sum to total due to rounding.		

The Columbia statuses with four percent or more of the hits on *Past Masters* were:

Columbia Status	**Hits**	**Share of Hits**
Undergraduate Student	86	69%
Graduate Student	11	9%
Unidentified	10	8%
Total	107	86%
Note: Detail may not sum to total due to rounding.		

4.3.4.3.2.2 Columbia University Press

The departments with four percent or more of the 597 hits on the Columbia University Press texts were:

Department	**Hits**	**Share of Hits**
Social Work	547	92%
Unidentified	28	5%
Total	575	96%
Note: Detail may not sum to total due to rounding.		

The Columbia statuses with four percent or more of the hits on these texts were:

Columbia Status	**Hits**	**Share of Hits**
Graduate Student	525	88%
Unidentified	28	5%
Faculty (Professor-Lecturer)	23	4%
Total	576	96%
Note: Detail may not sum to total due to rounding.		

4.3.4.3.2.3 Oxford University Press

The departments with four percent or more of the 110 hits on the Oxford University Press texts were:

Department	Hits	Share of Hits
International & Public Affairs	26	24%
Social Work	21	19%
Engineering	10	9%
Political Science	8	7%
Unidentified	8	7%
Physics	5	4%
Lamont-Doherty Observatory	5	4%
Total	83	75%
Note: Detail may not sum to total due to rounding.		

The Columbia statuses with four percent or more of the hits on these texts were:

Columbia Status	Hits	Share of Hits
Undergraduate Student	35	32%
Graduate Student	30	27%
Professional Student	26	24%
Unidentified	8	7%
GRA	5	4%
Total	104	95%
Note: Detail may not sum to total due to rounding.		

4.3.4.4 Online Book Use Per User

The distribution of number of hits on the online books collection per user over this period indicates that while many users are making quite cursory use of the online books, more are looking at more than one file (e.g., reference entry, chapter) in the collection.

Table 28. Distribution of Hits Per Unique User, March 15 - May 31, 1997

No. Of Hits Per User	% of Total Users
1	34%
2	16%
3	8%
4	8%
5	4%
6 - 10	16%
11 - 15	5%
16 - 20	5%
21 - 25	2%
>25	2%
Detail may not sum to 100% due to rounding.	

The distribution of number of unique titles viewed by these users over this period indicates that most users come to the collection to look at a single book. The greatest number of books used by a single person was seven (by two persons).

Table 29. Distribution of Unique Titles Viewed Per User, March 15 - May 31, 1997

No. Of Titles Viewed Per User	Number of Users	% of Total Users
1	225	80%
2	32	11%
3	11	4%
4	8	3%
5	1	*%
6	1	*%
7	2	1%
Total	280	100%
Note: Detail may not sum to total due to rounding. * Less than 0.5%		

Not surprisingly, there is a certain correlation between number of hits and number of titles used. Those with only one hit could only have looked at one title (42 percent of those using one book). The range of hits among those who used only one book is wide - 20 (9 percent) had more than ten hits. Six users had more than 25 hits; two of them looked at only one book, one each at two and three books, and two at seven books. These statistics indicate some significant use of the collection as measured by average number of hits per title used.

However, hits on several titles need not indicate heavy use of the online books collection. The individual who looked at five books had a total of only six to ten hits as did four of the seven people who looked at four books (one to two hits each). The person who looked at six books had 11 to 15 hits in total (an average of about two hits per book).

As the following table shows, graduate students tended to have more hits, undergraduates and faculty fewer hits.

Table 30. Hits Per Unique User by Academic Cohort, March 15 - May 31, 1997

Academic Cohort	N=	1 Hit	2-3 Hits	4-5 Hits	6-10 Hits	11-20 Hits	>20 Hits
Undergraduate	114	40%	28%	13%	14%	4%	1%
Grad. Student	66	18%	14%	9%	20%	27%	12%
Prof'l Student	9	33%	22%	22%	11%	0%	11%
Faculty	12	42%	25%	17%	8%	8%	0%

These are highlights of the recent data on usage by individuals. Once we have the information on sessions, we will be able to derive valuable information on user behavior - not only number of books used and hits on those books but parts of the book used and repeat usership. We will begin to be able to see revealed preference in user behavior and will be less reliant on responses to questionnaires.

……

Cover Page Table of Contents Next Page Footnotes

Previous Page

For additional information about the conference, or The Andrew W. Mellon Foundation's scholarly communication initiatives, please contact Richard Ekman. For additional information about ARL or this web site contact Patricia Brennan, ARL Program Officer at (202) 296-2296.

Return to Office of Scholarly Communication Home Page

ARL Home ARL Scholarly Communication and Technology Home Page

© Association of Research Libraries, Washington, DC
Web Design by Angelo F. Cruz
Maintained by ARL Web Administrator

Scholarly Communication and Technology

Conference Organized by The Andrew W. Mellon Foundation

at Emory University
April 24-25, 1997

Copyright © of the papers on this site are held by the individual authors or The Andrew W. Mellon Foundation. Permission is granted to reproduce and distribute copies of these works for nonprofit educational or library purposes, provided that the author, source, and copyright notice are included on each copy. For commercial use, please contact **Richard Ekman** at the The Andrew W. Mellon Foundation.

Session #4 Patterns of Usage

Online Books at Columbia: Measurement and Early Results on Use, Satisfaction, and Effect

Carol A. Mandel
Deputy University Librarian
Columbia University

and

Mary C. Summerfield
Coordinator, Online Books Project
Columbia University Libraries

and

Paul Kantor
Consultant

……

5. UNDERSTANDING FACTORS INFLUENCING USE

One of our principal goals is to understand what distinguishes users from non-users, early adopters from late adopters, and so on. We have tentatively identified several kinds of variables that might influence users with regard to adoption of electronic books. We can broadly divide these into: resource factors (discussed here), attitude factors (discussed in section 6), and behavior factors (discussed in section 7).

5.1 Resource Factors

We believe the key resource factor to be possession of, or easy access to, adequate computer equipment and connections, so that online books are a reasonable alternative to paper versions. As noted earlier in a discussion of early research results, we have two primary methods of studying this factor:

1. Asking the question about access to computers with network connections that was discussed earlier; and

2. Developing a detailed profile of the computer resources available to representative samples of members of the Columbia community. As described earlier, we used the campus computing survey instrument in three sweeps through the community, in what we regard as pilot implementations in Fall 1995 and Winter 1996. The questionnaire asked respondents for detailed information about their computers. However, many respondents did not supply the requested data on computer power, size of hard drive, modem speed, and the like. Furthermore, observation of developments in the personal computer marketplace makes it clear that changes in personal computers will make the top of the original scales on these questionnaires the bottom of the scales in as little as two years.

Section 7 contains further analyses of the first of these factors.

5.2 Attitude Factors - In-Class Survey

We cannot probe attitude factors easily in a simple survey, whether in paper or online. We have designed some questions aimed at assessing whether the respondent thinks that members of his peer group use and/or prefer electronic access to books and other resources. On the one hand, this perception of others' preferences might precede and shape a user's own preferences and behavior. Alternatively, if use of computer modalities is, as some psychological research suggests, a very private activity, awareness of the behavior of others may, in fact, lag the move to using online books.

5.2.1 The In-Class Survey

Examples of such results include the data from Fall 1996 and Spring 1997 in-class surveys of students who were assigned a reading that was available online for the session during which the survey was administered. Students who had done the reading were asked to answer questions about what forms of the book they used, how long they used that form, where they did that studying, format preference, reasons for it and impacts, expectations for format most used by classmates, and the two benchmark questions about computer access and time online. (See Exhibit 4.) Those who had not done the assignment were asked to respond only to the last three questions.

Exhibit 4. Survey Of Book Use For Course Readings

As part of its effort to serve you better, the Libraries would like to know what methods you used in reading an assignment for this class session. All responses will be kept confidential.

A. Did you read the assignment in Kadushin's **SUPERVISION IN SOCIAL WORK** for this class session?
1. YES *(If so, please answer all the questions.)* 2. NO (*If not, please skip down and answer Questions F-I* **only.)**

B. Following is a list of methods that you might have used in doing this reading. Please tell us about your use of each for this assignment. If you used a method, please tell us for about how long you used it and where you did this reading.

Methods of Reading This Assignment	**Did you use it? (Please circle)**	**For about how long? # Hours & # Minutes**		**Where (e.g., library, dorm room, lounge, classroom)?**
1. Your own copy of the book	1.YES 2.NO			
2. A friend's copy of the book	1.YES 2.NO			
3. A library copy of the book	1.YES 2.NO			
4. Photocopy from paper copy	1.YES 2.NO			
Using CWeb Online Text:				
5. Reading it directly from CWeb	1.YES 2.NO			
6. JAKE printout of text	1.YES 2.NO			
7. Printout using non-JAKE printer	1.YES 2.NO			
8. Download of online text to disk & reading away from CWeb	1.YES 2.NO			

C. If you used more than one method, which one did you like best? (Please circle the number of the preferred method from the above table.) **1 2 3 4 5 6 7 8**

D. Why did you like that method best? (Please circle the numbers of all the reasons that apply.)
1. Less costly 2. Easy to get to 3. Easy to read 4. Always available 5. Easy to copy 6. Easy to search for words or concepts 7. Easy to annotate/take notes 8. Other reasons:

E. What were the impacts on your work of using the method you liked best? (Please circle the numbers of all that apply.) 1. I learned better. 2. I learned faster. 3. Learning was more fun. 4. I was more likely to do the assignment. 5. Reading the assignment was more difficult. 6. Doing the assignment was faster. 7. Doing the assignment was slower. 8. Other impacts:

F. Which of these methods of reading this assignment do you think was most used by your classmates? (Please circle the number of the method from the above table.) **1 2 3 4 5 6 7 8**

G. Is there a computer connected to the campus network (by modem or direct link) that you can use whenever you want? (Please circle.) 1. YES 2. NO

H. About how many ***hours per week*** do you spend in each of the following online activities?

_____ Email _____ Listservs and Newsgroups ______ CLIO Plus

_____ Scholarly Text, Image or Numeric Data Sources ______ Other WWWeb

I. Your insights into your experience and preferences in using various book formats are valuable:

Thank you for your assistance with this study.

Form 8, 9/96: Distribution Date: 4/29/97 Course: SOCW T7134

In Fall 1996, most of this surveying (67 percent of a total of 439 responses) was done in sections of Columbia College's Contemporary Civilization course for which some of the readings are available in the *Past Masters* set of humanities texts. However, those students are expected to use the assigned editions of the readings and to bring a copy to class for use during the discussion. This may well have biased students' choices in methods of reading the assignments. About 16 percent of the cases in the sample came from graduate Social Work classes and 17 percent from a large upper level undergraduate political theory class.

In Spring 1997, the Contemporary Civilization course was the source of 106 (44 percent) of the 239 respondents. Two political science courses, one undergraduate and one graduate level, were the source of another 17 respondents (7 percent). Four graduate Social Work courses were the source of the remaining 116 respondents (49 percent).

5.2.2 Methods of Studying A Class Reading

Some students had not done the reading for the course session in which the surveying was done. Some others reported using more than one method.

Table 31. Methods of Reading This Assignment: Whole Sample Fall 1996 And Spring 1997

	Fall 1996		**Spring 1997**	
Methods of Reading This Assignment	**Count**	**% of Responses**	**Count**	**% of Responses**
Used Own Copy	269	70%	141	73%
Used Friend's Copy	54	14%	20	10%
Used Library Copy	33	8%	17	9%
Used Photocopy	11	3%	17	9%
Reading it directly from CWeb	0	0%	0	0%
JAKE printout of text	10	3%	16	8%
Printout using non-JAKE printer	4	1%	4	2%
Download of online text to disk & reading away from CWeb	5	1%	1	*%
Total	**386**	**100%**	216	**100%**

This table shows that, in Fall 1996, 70 percent of the responses reported using one's own copy of the text. The next most common method was to use a friend's copy (14%). The shares for those two modes are insignificantly different in Spring 1997.

The questionnaire gives four alternative means of *Using CWeb Online Text:*

Reading it directly from CWeb

JAKE printout of text

Printout using non-JAKE printer

Download of online text to disk & reading away from CWeb

In Fall 1996, there were 19 reports (5 percent) of printing out or downloading from the CWeb books, but none of reading directly from CWeb. In 89 percent of those cases, the respondent was not in a Contemporary Civilization class. In Spring 1997, about 11 percent of responses reported using some form of the online text, but again none reported reading on screen.

5.2.3 Preferences for Studying Class Reading

There were far fewer responses (119 in Fall 1996 and 88 in Spring 1997) as to the preferred mode of studying. In both semesters about two thirds of respondents reported that reading their own copy was preferred.

Table 32. Preferred Method Reading This Assignment: Whole Sample Fall 1996 and Spring 1997

	Fall 1996		Spring 1997	
Preferred Method of Reading Assignment	**Count**	**% of Cases**	**Count**	**% of Cases**
Own Copy	83	67%	56	64%
Friend's Copy	9	8%	6	7%
Library Copy	10	8%	6	7%
Photocopy	7	6%	8	9%
Reading it directly from CWeb	2	2%	7	8%
JAKE printout of text	7	6%	6	7%
Printout using non-JAKE printer	3	2%	5	6%
Download of online text to disk & reading away from CWeb	3	2%	1	1%
Total Responses	124	101%	95	109%
Total Cases Responding	**119**		**88**	
Note: Detail may not sum to 100% due to rounding.				

5.2.4 Reasons for Preference

As the following table shows, in both semesters, the three strongest reasons for preference were *always available*, *easy to annotate,* and *easy to read*, with the last two reasons switching position between the semesters.

Table 33. Reason for Preferred Method: Whole Sample Fall 1996 and Spring 1997

	Fall 1996		Spring 1997	
Reasons for Preference	**Count**	**% of Cases**	**Count**	**% of Cases**
Less Costly	60	22%	33	23%
Easy to get to	0	0%	0	0%
Easy to Read	104	38%	70	48%
Always Available	199	72%	108	74%
Easy to Copy	21	8%	20	14%
Easy to Search for Words	30	11%	15	10%
Easy to Annotate	135	49%	57	39%
Other Reasons	25	9%	16	11%
Total Responses	574	209%	319	219%
Total Cases Responding	**276**		**146**	
Note: Respondents could give more than one reason for their preference.				

At present, these attributes are possessed only by a personal copy or photocopies from print copies or printouts from electronic copies. (Online books are *always available*, but one assumes that ready physical access to a computer does not meet the criterion *always* as students interpreted it here.)

The cross-tabulation of preferred method of use and reasons for that preference produces logically consistent results. For example, all of the respondents who gave *Printout using non-JAKE printer* or *Download of online text to disk and reading away from CWeb* as their preferred method gave *less costly* as one of their reasons while few of the those preferring their own copy gave that reason.

Table 34. Preferred Method & Reason for Preferred Method: Whole Sample Spring 1997 - Row Percentages

	Reason for Preference						
Preferred Method	**Less Costly**	**Easy to Read**	**Always Available**	**Easy to Copy**	**Easy to Search**	**Easy to Annotate**	**Other**
Own Copy	8%	47%	83%	7%	9%	41%	11%
Friend's Copy	75%	25%	75%	0%	0%	25%	0%
Library Copy	25%	25%	50%	25%	25%	0%	25%
Photocopy	50%	50%	33%	33%	0%	17%	0%
On CWeb	60%	20%	80%	20%	40%	0%	0%
JAKE Printout	50%	67%	83%	83%	0%	17%	0%
Other Printout	100%	20%	60%	60%	20%	20%	0%
Download to Disk	100%	0%	100%	100%	0%	0%	0%
Total	24%	44%	75%	16%	12%	33%	9%
Note: *Easy to Get To* was another reason offered but no one chose it in either semester.							

So we have a consistent picture of what makes a mode good, preferred, and used. In other words, our student respondents are behaving rationally.

5.2.5 Impact of Preferred Method

As the following table shows, when asked what the impact of the various possible modes was, a majority of the students selected *more likely to do the assignment. Learned better* and *doing assignment faster* ranked second and third, being cited by about one third of the students.

Table 35. Nature of Impact of Preferred Method: Whole Sample Fall 1996 and Spring 1997

	Fall 1996		Spring 1997	
Impact of Preferred Method	**Count**	**% of Cases**	**Count**	**% of Cases**
Learned Better	92	34%	51	34%
Learned Faster	51	19%	27	18%
Learning More Fun	16	6%	13	9%
More Likely To Do Assignment	154	57%	82	54%
Reading More Difficult	8	3%	1	1%
Doing Assignment Faster	85	31%	56	37%
Doing Assignment Slower	12	4%	2	1%
Other Impacts	29	11%	21	14%
Total Responses	**447**	**165%**	253	168%
Total Cases Responding	**271**		**152**	
Note: Detail may not sum to 100% due to rounding.				

These are again entirely rational bases for preferring some particular mode, when we note that a student's role is to get assignments done and to learn.

5.2.6 Comparison of Personal Behavior and Perception of Others' Behavior

Students were asked what method they thought most of their classmates used in order to learn whether they perceived a shift to using the online materials. In Fall 1996, 81 percent of respondents chose *own copy*, and 14 percent chose *library copy*. In Spring 1997, these values were 80 percent and eight percent, respectively. This contrasts with the reality that 70 percent (73 percent in Spring 1997) used their own copies, 14 percent (ten percent) used a friend's copy and eight percent (nine percent) used a library copy.

Interestingly, while none of the respondents had used the CWeb book directly to do this assignment, in both survey periods almost two percent (five or six students) gave that response to this question. In the fall, another nine students (two percent) and, in the spring, another 13 students (six percent) thought that their classmates had used some form of print copy or downloaded file from CWeb. Thus, students are over-estimating their colleagues' propensity to read directly from CWeb and under-estimating their propensity to read printed copy and downloaded files from CWeb.

Table 36. In-Class Surveys: Personal Behavior and Perception of Methods Used by Classmates to Read This Assignment: Whole Sample Fall 1996 and Spring 1997

	Fall 1996		**Spring 1997**	
Methods of Reading This Assignment	***Own Behavior***	***Perception of Others' Behavior***	***Own Behavior***	***Perception of Others' Behavior***
Used Own Copy	74%	81%	73%	80%
Used Friend's Copy	15%	7%	10%	4%
Used Library Copy	9%	14%	9%	8%
Used Photocopy	3%	4%	9%	7%
CWeb Directly	0%	2%	0%	2%
Used JAKE Print Copy	3%	1%	8%	5%
Used Other Print Copy	1%	*%	2%	1%
Used Download Copy	1%	1%	*%	*%
Note: * Less than .5%.				

At the present time, there is not sufficient penetration of the market by the online modes for us to draw any meaningful conclusions about leading and lagging impacts. Results of this survey in Fall 1997, particularly in Social Work classes, should give some indication of trends as students will have had more time in which to gain awareness of the availability and attributes of the online format.

* * *

Cover Page Table of Contents Next Page Footnotes

Previous Page

For additional information about the conference, or The Andrew W. Mellon Foundation's scholarly communication initiatives, please contact Richard Ekman. For additional information about ARL or this web site contact Patricia Brennan, ARL Program Officer at (202) 296-2296.

Return to Office of Scholarly Communication Home Page

ARL Home ARL Scholarly Communication and Technology Home Page

© Association of Research Libraries, Washington, DC

Web Design by Angelo F. Cruz

Scholarly Communication and Technology

Conference Organized by The Andrew W. Mellon Foundation

at Emory University
April 24-25, 1997

Copyright © of the papers on this site are held by the individual authors or The Andrew W. Mellon Foundation. Permission is granted to reproduce and distribute copies of these works for nonprofit educational or library purposes, provided that the author, source, and copyright notice are included on each copy. For commercial use, please contact **Richard Ekman** at the The Andrew W. Mellon Foundation.

Session #4 Patterns of Usage

Online Books at Columbia: Measurement and Early Results on Use, Satisfaction, and Effect

Carol A. Mandel
Deputy University Librarian
Columbia University

and

Mary C. Summerfield
Coordinator, Online Books Project
Columbia University Libraries

and

Paul Kantor
Consultant

……

6. THE ONLINE SURVEY

6.1 The Online Interview Instrument

The online instrument is mounted as an HTML form. The key questions are presented here along with an example of the pull down list that accompanies one of the questions.Exhibit 5. Online Survey Instrument: Non-Reference Books

---

A. What is the title of the book you just used?

B. Please select the best description of that work or project for which you are using this book. If "Other", please specify:

1='Research project, e.g., paper, book'

2='Class preparation'

3='Current awareness in field'

4='Other University activity'

5='Other:

C. How long ago did you recognize the need to consult this book for this use?

__ [weeks__]

D. How soon do you expect to make use of what you get from this book?

__ [weeks__]

E. What did you do with this book on this occasion? (Select all applicable uses.):

__ Looked up something etc.

F. Which forms of this book have you ever used? (Select all that apply by checking the check boxes in the left column [Used] below.) If you have used this book in more than one way, which one do you prefer overall? (Select one of the 'radio' buttons in the right column [Prefer] below.)....

G. Referring to the way of using this book that you prefer, why do you like it best? (Select all that apply.)....

H. On how many occasions (including this one) have you used this book in any format during the last 3 months?

I. For approximately how many minutes in total have you used this book during the last 3 months?

J. About how many times in the past 12 months have you used an online book, i.e., a monograph or reference book available on CNet or another computer network? __ times.

K. In the type of work you are doing now, do you find that paper books or online books help you to be more productive?

L. Do you find that you are able to produce results of higher quality when you use paper books or online books?

M. Is there a computer attached to the campus network (by modem or direct link) that you can use whenever you want? Yes / No (*)

N. About how many hours per week do you spend in each of the following online activities?

Email: __ Listservs & Newsgroups: __ CLIO-Plus: __

Text/Image/Numeric Data Sources on WWW: __ Other WWW:__

O. What is your present primary relationship to Columbia?

[Undergraduate______] If "Other", please specify:

P. What is your primary discipline?

[Undetermined__________________] If "Other", please specify:

We initially launched the Web questionnaire in two parts. The reader was given the initial part, which asked questions that could be answered before he used the book, e.g., about his reason for using the book, timing of need for the material he was seeking, and his status, when he clicked on the title of the book. He was not required to complete it in order to move on to the book, but it was easy to respond at least in part. The scholar was asked to click on the button taking him to the second part when he finished his session with the book; it asked various questions about how he felt about the online format. We could not force the user to go to the second questionnaire and hardly anyone did. At the same time, the online book designers found working with two questionnaires to be difficult.

In preparation for the Fall 1996 semester, we switched to a single questionnaire format in which the scholar must choose to go to the questionnaire after he uses the book. Response rates have been poor with fewer than ten questionnaires submitted in any week and many of those responses incomplete.

Data captured from the questionnaires are processed (using Unix utilities) to produce a standard data file for input into SAS or SPSS. Findings from the most recent data are summarized below.

6.2 Early Results from CWeb Survey

6.2.1 CWeb Survey Responses by Online Text Used

From late September 1996 through early June 1997, we received 85 responses to the CWeb questionnaire.

Table 37. CWeb Online Survey Responses by Online Text Used, September 1996 - June 1997

Online Text Used	Count	% of Total
Oxford English Dictionary	64	75%
Granger's Index to Poetry	1	1%
Garland Reference Works	2	2%
Past Masters Texts	8	9%
CUP Social Work	7	8%
Other CUP Monographs	0	0%
OUP Monographs	2	2%

The OED is both the most used of the online books and the one for which the most survey responses were returned.

Given *The OED*'s overwhelming presence in the responses, this analysis is largely one of reactions to the online *OED*. In a few cases the analysis distinguishes between *The OED* and all of the other texts.

6.2.2 CWeb Survey: Primary Project for Using Book

The questionnaire asked the scholar to select the *best description of that work or project for which you are using this book* and gave a choice of five options. The distribution of responses was:

Table 38. CWeb Online Survey Responses by Work Involved, September 1996 -

Table 38. CWeb Online Survey Responses by Work Involved, September 1996 - June 1997

Work/Project	% of Responses
Research project	46%
Class preparation	28%
Current awareness	6%
Other University activity	5%
Other	15%

Research projects are the major purpose for using the online books.

6.2.3 CWeb Survey: Ways of Using Book

The questionnaire asks *What did you do with this book on this occasion? (Select all applicable uses.)* It offers different reasons for the monographs and for the various reference books.

For the *OED* responses, the distribution of book uses was:

Table 39. CWeb Online Survey Responses: Uses of *The OED* , September 1996 - June 1997

Use	% of OED Responses
Definitions	94%
Etymology	43%
Pronunciation	14%
History of words	44%
Examples of Use	36%
Citations for authors	13%
Citations for eras	21%

For all the other books, the distribution of uses was:

Table 40. CWeb Online Survey Responses: Uses of Other Books, September 1996 - June 1997

Use	% of Other Responses
Looked up something	32%
Searched for something	47%
Looked at citations	16%

BEST COPY AVAILABLE

Looked at table of contents &/or index	11%
Looked at introduction &/or conclusions	26%
Looked at graphics	16%
Read part of the book	68%

Those who reported that they used the online book by reading part of it were asked how much they read. Responses were distributed as follows:

Less than 10%	59%
10-30%	18%
Over 30%	23%

The majority of these online book users read less than 10 percent, say one chapter, online.

6.2.4 CWeb Survey: Forms of This Book Ever Used

The questionnaire asks *Which forms of this book have you ever used?* and offers the scholar nine options. Responses were distributed as follows:

Table 41. CWeb Online Survey Responses: Forms of the Book Ever Used, September 1996 - June 1997

Forms Ever Used	% of OED Responses (N=60)	% of Other Responses (N=17)
Online copy in library	43%	12%
Online copy elsewhere	53%	59%
Printout from online copy	25%	29%
Download from online	13%	12%
Library paper copy	60%	12%
My own paper copy	30%	24%
Colleague's paper copy	13%	12%
Photocopy from paper copy	10%	0%
CD-ROM	12%	6%

For *The OED, paper copy in the library* received the most mentions with *online copy elsewhere* coming in a close second. For the other books, *online copy elsewhere* was the dominant response.

6.2.5 CWeb Survey: Preferred Form of This Book

The questionnaire asked *If you have used this book in more than one way, which one do you prefer overall?* The same choices were offered as above. Responses (56 for *The OED* and 17 for the other books) were distributed as follows:

Table 42. CWeb Online Survey Responses: Preferred Book Form, September 1996 - June 1997

Preferred Form	% of OED Responses	% of Other Responses
Online copy in library	21%	12%
Online copy elsewhere	46%	24%
Printout from online copy	12%	24%
Download from online	2%	6%
Library paper copy	9%	0%
My own paper copy	5%	35%
Colleague's paper copy	0%	0%
Photocopy from paper copy	0%	0%
CD-ROM	4%	0%

Online copy used outside the library is far the preferred book form for *The OED* with more than twice the votes as the next most preferred form, *online copy used in the library. Printout from online copy* ranked third. The various forms of using the online *OED* received over 80 percent of the preferences votes.

The responses for the other books are also revealing. Just over a third of respondents preferred *my own paper copy*. Given the attributes ranked as important - always available and easily annotated, this is a logical top runner for non-reference books. However, various forms of using the online book received all the other votes of this small sample of users of the online book collection.

6.2.6 CWeb Survey: Reasons for Preference

The questionnaire asked *Referring to the way of using this book that you prefer, why do you like it best? (Select all that apply.)* Responses were distributed among the options offered as follows:

Table 43. CWeb Online Survey Responses: Reasons for Book Form Preference, September 1996 - June 1997

Reasons for Preference	% of OED Responses	% of Other Responses
Less costly	41%	60%
Easy to get to	71%	75%
Easy to read	49%	40%
Always available	66%	75%
Easy to search	73%	40%
Easy to copy	44%	30%
Easy to take notes/annotate	20%	30%
Other reasons	8%	5%

Easy to get to, which had no mentions in the in-class survey, was the most popular response given in this survey. In part this reflects the heavy presence of *The OED* in this survey, but this reason for the preference also tied with *always available* for the other responses involving books other than *The OED*.

6.2.7 CWeb Survey: Preferred Format and Reasons for Preference

Looking at all the responses, the top reasons for each format being preferred were:

Table 44. CWeb Online Survey Responses: Preferred Book Form and Key Reasons for Preference, September 1996 - June 1997

Preferred Form	Key Reasons for Form Preference	
Online copy in library	Easy to get to	Easy to search
Online copy elsewhere	Easy to get to	Always available
Printout from online copy	Always available	Less costly Easy to get to Easy to copy
Download from online	One mention for all but *Easy to read*	
Library paper copy	Easy to get to Always available	Easy to read Easy to search
My own paper copy	Always available	Easy to read
Colleague's paper copy	Not preferred	
Photocopy from paper copy	Not preferred	
CD-ROM	Easy to search	Easy to copy

The popularity of *easy to get to* is consistent with the preference for *online copy used outside the library* as long as the respondent has easy access to a computer with a Web browser.

6.2.8 CWeb Survey: Frequency of Use in Past Three Months

The questionnaire asked On how many occasions (including this one) have you used this book in any format during the last 3 months? The 79 responses were distributed as:

Table 45. CWeb Online Survey Responses: Frequency of Use of This Book in Past three Months, September 1996 - June 1997

Number of Occasions	Number of Responses	% of Responses
0	6	8%
1	16	20%
2	12	15%
3-4	10	13%
5-6	10	13%
7-8	4	5%
10-12	9	11%
15-19	3	4%
20-35	6	8%
50-99	3	4%

Those responding 'zero' were not following the directions to the question and presumably meant that this was their first occasion to use this book in this period. The mean was 8.7 occasions and the median 3.0 occasions - or an average of about three occasions per month based on the mean or once a month based on the median. It may be that heavy users of online books are more likely to notice our questionnaire and ultimately to respond and, hence, to be over-represented in this sample. However, the question asks about use in all formats.

6.2.9 CWeb Survey: Total Usage in Minutes in Past Three Months

The questionnaire asked *For approximately how many minutes in total have you used this book during the last 3 months?* The 79 responses were distributed as:

Table 46. CWeb Online Survey Responses: Total Usage of This Book In Minutes in Past Three Months, September 1996 - June 1997

Minutes	Number of Responses	% of Responses
0	2	2%
1-9	13	16%

10-12	16	20%
15-18	11	14%
20-24	10	13%
25-36	15	19%
45-60	10	13%
80-90	2	2%

Again, those responding 'zero' were not following the directions to the question and presumably meant that they had not spent any time with this book previously in this period. The mean was 22 minutes and the median 15 minutes. These are not great amounts of time for using a monograph but they are substantial for using a dictionary.

6.2.10 CWeb Survey: Frequency of Use of Any Online Book in Past Year

The questionnaire asked *About how many times in the past 12 months have you used an online book, i.e., a monograph or reference book available on CNet or another computer network?* The 75 responses were distributed as:

Table 47. CWeb Online Survey Responses: Total Usage of Online Books In Past Year, September 1996 - June 1997

Number of Times	Number of Responses	% of Responses
0	10	13%
1-2	18	24%
3-6	15	20%
10-16	8	11%
20-25	9	12%
30-50	12	16%
75-99	3	4%

The mean was 15 uses in the past year and the median five uses. This sample is most likely not representative of all users of the online books, let alone of the Columbia community.

6.2.11 CWeb Survey: Effect of Online Books on Scholarly Work

Two key questions asked on all of our questionnaires, other than those distributed in class, seek to determine the effect of online books on scholarly work.

- *In doing the type of work for which you used this book, do paper books or online books help you be more productive?*
- *Do you find that you are able to do work of higher quality when you use paper books or online books?*

The questionnaire offers a range of seven responses from *Much greater productivity (quality) with paper* through *No Difference* to *Much greater productivity (quality) with online* plus *Cannot Say.*

6.2.11.1 CWeb Online Survey: Productivity, Book Type and Format

As the following table shows, many *OED* users felt that they are more productive using the online *OED* works while only a modest number of the users of the other online books felt that they are more productive using online books.

Table 48. CWeb Online Survey: ***In doing the type of work for which you used this book, do paper books or online books help you be more productive*****? by Book, September 1996 - June 1997**

Response	OED (N=64)	All Other Books (N=21)
Cannot Say	12%	10%
Paper Much Greater	16%	24%
Paper Greater	8%	14%
Paper Somewhat Greater	12%	14%
No Difference	2%	19%
Online Somewhat Greater	17%	5%
Online Greater	17%	5%
Online Much Greater	16%	10%
Note: Detail may not sum to 100% due to rounding.		

Of the group of 64 users of *The OED*, 50 percent believed that they were more productive with *online books* and 36 percent believed that they were more productive with *print books*. Only one respondent thought there was no difference and eight responded *cannot say.* The 21 users of the other books did not share this feeling. Only 19 percent believed that they were more productive with *online books* and 48 percent believed that they were more productive with *print books*. However, another 19 percent noted *no difference* in productivity and ten percent responded *cannot say.*

6.2.11.2 CWeb Online Survey: Work Quality, Book Format and Type

As the following table shows, the distribution of responses to the second question about the quality of work when using print and online books supports the print format in general, although many respondents found no difference in their work quality with the two formats.

Table 49. CWeb Online Survey: ***Do you find that you are able to do work of higher quality when you use paper books or online books*****? by Book, September 1996 - June 1997**

Response	OED (N=64)	All Other Books (N=21)
Cannot Say	16%	5%
Paper Much Greater	16%	24%
Paper Greater	6%	14%
Paper Somewhat Greater	16%	14%
No Difference	31%	29%
Online Somewhat Greater	2%	0%
Online Greater	8%	0%
Online Much Greater	6%	14%
Note: Detail may not sum to 100% due to rounding.		

For *The OED*, 37 percent supported print books, 16 percent backed online books, and 31 percent perceived no difference in work quality. For all other books, 52 percent voted for print books, 14 percent for online books, and 29 percent perceived no difference in quality.

These responses are somewhat puzzling as the reference book most used online is *The OED* and the features of the CWeb version provide as much utility if not more than the print version (with the exception of being able to view neighboring entries at a glance).

Cross-tabulation of these two questions finds considerable correlation in the responses - those who supported the paper version for productivity tended to support it for quality as well.

Table 50. CWeb Online Survey: Quality and Productivity, September 1996 - June 1997

	Quality of Work			
Productivity	**Cannot Say**	**Better Paper**	**No Difference**	**Better Online**
Cannot Say	8	0	2	0
Better Paper	3	27	5	1
No Difference	0	1	4	0
Better Online	0	9	15	12

Almost a third of the 85 respondents ranked paper books as yielding both greater productivity and greater quality, while only one person ranked paper books better for productivity and online books better for quality. About 14 percent ranked online books better on both scores, while about ten percent ranked online books better for productivity but paper books better for quality.

6.2.12 CWeb Online Survey: Columbia Cohort of Respondents

The questionnaire asked a respondent to select one of several statuses offered as that which represented his present primary relationship to Columbia University. The responses were distributed as follows.

Table 51. CWeb Online Survey: Respondent's Columbia Status, September 1996 - June 1997

Columbia Status	Number of Responses	% of Responses
Undergraduate	49	58%
Graduate Student	20	24%
Faculty	6	7%
Non-Faculty Officer	3	4%
Staff	5	6%
Special Student	1	1%
Other	1	1%

Of the 85 respondents whose questionnaires were analyzed above, 58 percent were undergraduates and 24 percent graduate students. This is consistent with the server data on *OED* user status, which identified 58 percent of users and 55 percent of hits with undergraduates, and six percent of users and 11 percent of hits with graduate students.

6.2.13 CWeb Online Survey: Discipline of Respondents

The questionnaire asked a respondent to select one of 16 disciplines (including *Other*) as that which defined his scholarly focus. The 85 responses were distributed as follows.

Table 52. CWeb Online Survey: Respondent's Discipline, September 1996 - June 1997

Discipline	Number of Responses	% of Responses
Undetermined	44	52%
Architecture	1	1%
Art	3	4%
Business	2	2%
Computer Science	9	11%
Engineering	4	5%
Health Sciences	6	7%
History	2	2%
Humanities	14	16%

As the table shows, as might be expected, many of the undergraduate respondents have not yet selected a discipline. There were no representatives of seven possible disciplines, including major ones such as Social Work, Social Sciences, and Natural and Physical Sciences, in the responses.

6.2.14 Online Survey: Place in Project

We will need to explore the responses to this survey closely now and as we track it in the future and utilize our findings in structuring the interviews we undertake in the months ahead. It would be surprising if we do not see a shift in responses as our collection grows and as users have an opportunity for continuing use. Of course, we may have difficulty eliciting repeat responses to our questionnaire from the same individuals. However, perhaps some repeat users who have not completed the questionnaire will do so in the future. If necessary, we will be more aggressive in seeking feedback from users, e.g., by sending them questionnaires or interview requests in email or by telephone.

We are exploring various methods to increase our response rate. From March 15 to May 31, 1997, there were 42 hits on the survey button; 14 on the *OED* survey and 28 on the monograph survey. In this period, 280 people used the online book collection; thus, only 15 percent of them went to the survey during any of their sessions with the collection. During this period, 22 completed surveys were submitted, for a 52 percent return on surveys viewed. We are hopeful that introducing a frames design to our books, with the survey button on the frame along with navigational and search buttons, will remind users about the survey and encourage them to go to it and complete it. Clearly, getting that initial interest is critical to getting users to assist with our research by completing the questionnaire.

Other options we are exploring include breaking up the online questionnaire so that users confront only a screen full of questions (i.e., each respondent would answer only a subset of our questions), however, the non-response problem is one of getting the users of the online books to click on the questionnaire button much more than one of getting them to complete the questionnaire once they have done that. We are exploring changing our incentives, such as by instituting an improved lottery, but changes to date have not had a notable impact.

6.3 User Comments

We are gathering more contextual feedback from users through follow-up questions on email and through personal interviews. We have been using this feedback in making design decisions and we will be pulling it together more systematically over the course of this semester and early next summer.

Comments on questionnaires help us keep grounded in our work. The following example, quoted in full, shows remarkable insight into the complexities of assessing impact in a rapidly changing environment. It was anonymous.

Exhibit 6. An Extended Comment

> *Your questions show a decided bias that attempts to lead the technology-shy into giving a negative review. You already know that this is a better method of text distribution! Why is this survey even here? There are only 2 advantages that books could possibly have over online texts. 1. They are easier to read. That issue will shortly become moot as people simply become accustomed to reading the texts on a screen rather than on a page. 2. They are portable. Online sources are infinitely more portable in an abstract sense since they can be distributed swiftly all over the world. Physically, every computer terminal is a potential source. It won't necessitate everyone getting a laptop to make e-texts as portable as physical books, although that is happening. In sum, get with it! What the [expletive deleted] are you doing? You KNOW that even if people aren't using this resource fervently now, they will in 2-3 years! Get off your butts and start putting more texts online instead of writing inane, technophobic, leading polls.*

Other, more courteous responses call attention to the need for excellent search and browsing capabilities in online books. Some were praising the current design for its provision of these capabilities. Others were suggesting that better capabilities were needed. Users would particularly like to see more analytical tools in the CWeb *OED*. Analysis of these comments along with those made in the ongoing interviewing of users will come in the next stage of our reporting.

……

Cover Page Table of Contents Next Page Footnotes

Previous Page

For additional information about the conference, or The Andrew W. Mellon Foundation's scholarly communication initiatives, please contact Richard Ekman. For additional information about ARL or this web site contact Patricia Brennan, ARL Program Officer at (202) 296-2296.

Return to Office of Scholarly Communication Home Page

ARL Home ARL Scholarly Communication and Technology Home Page

© Association of Research Libraries, Washington, DC
Web Design by Angelo F. Cruz
Maintained by ARL Web Administrator
Last Modified: September 8, 1997

Scholarly Communication and Technology

Conference Organized by The Andrew W. Mellon Foundation

at Emory University
April 24-25, 1997

Copyright © of the papers on this site are held by the individual authors or The Andrew W. Mellon Foundation. Permission is granted to reproduce and distribute copies of these works for nonprofit educational or library purposes, provided that the author, source, and copyright notice are included on each copy. For commercial use, please contact **Richard Ekman** at the The Andrew W. Mellon Foundation.

Session #4 Patterns of Usage

Online Books at Columbia: Measurement and Early Results on Use, Satisfaction, and Effect

Carol A. Mandel
Deputy University Librarian
Columbia University

and

Mary C. Summerfield
Coordinator, Online Books Project
Columbia University Libraries

and

Paul Kantor
Consultant

……

7. BEHAVIORAL FACTORS

7.1 Access to Networked Computer

As noted earlier (see section 4.2), in all of the Project's surveys the following question is asked: *Is there a computer (in the library or elsewhere) attached to the campus network (directly or by modem) that you can use whenever you want?* Our hypothesis is that the easier a Columbia scholar's access to the campus network and materials on CNet and CWeb the more likely he is to adopt online resources, including the collection of online books. In addition, we want to track this measure over time to see how it changes. The responses to this question in the in-class and CWeb online surveys are summarized here.

An overwhelming majority (80 or 94 percent) of the 85 respondents to the CWeb online survey responded to this question in the affirmative.

The students responding to the in-class survey did not see themselves as having such easy access to a networked computer. Of the 239 who responded to this question, 68 percent answered in the affirmative. The percent responding in the affirmative for the different types of classes was:

Table 53. In-Class Surveys: Ready Access to Networked Computer: Whole Sample Spring 1997

Class Type	**N**	**% Responding *Yes***
Contemporary Civilization	102	82%
Graduate Political Science	5	20%
Undergraduate Political Science	12	75%
Social Work Masters Students	104	65%
Total Respondents	239	68%

As in the on-site library survey, undergraduates claim greater access to networked computers than masters students do.

Table 54. In-Class Survey: Preferred Method of Reading This Assignment and Access to Networked Computer: Whole Sample Spring 1997

Preferred Method of Reading Assignment	**Access to Networked Computer?**	
	Yes (N=62)	**No (N=23)**
Own Copy	66%	61%
Friend's Copy	6%	9%
Library Copy	6%	4%
Photocopy	6%	13%
Reading it directly from CWeb	11%	0%
JAKE printout of text	8%	4%
Printout using non-JAKE printer	5%	9%
Download of online text to disk & reading away from CWeb	2%	0%

As the above table shows, comparing students' perceived access to networked computers and their preferred book form reveals that such access does not lead students to preferring online books. Given the stated reasons for their preferences, this is logical. Over 66 percent of those responding Yes to this question preferred their own copy of a book while only 61 percent of those responding *No* did. Photocopy was the preferred form for 13 percent of those responding *No*, but for only six percent of those responding *Yes*. This combination of responses suggests that an economic element is at work here. Those who cannot afford their own computers may also prefer not to buy books for classes.

7.2 Time In Online Activities

Based on study of the data, we have settled on collection of information on the amount of time spent per week in various online activities to represent the behavior of the users. (This question was discussed earlier in the context of the on-site survey of library users.) The balance among the various online activities will vary with discipline, and with the position of the user. In the versions of the questionnaire in use since last spring for books in print and online format, the data have been gathered by the following question.

About how many hours per week do you spend in each of the following online activities?

Email: __ Listservs & Newsgroups: __ CLIO-Plus: __

Text/Image/Numeric Data Sources on WWW: __ Other WWW:__

Since these activities are all measured in hours per week, we can sum them to produce a single simple measure of the level of *online activity* (as we asked the question in the on-site library user survey). The results are instructive. We have prepared tables showing the percent distribution of respondents by number of hours spent online, in all activities, for three different groups: users of online books, users of the *OED* in paper format, and students surveyed in class. (Table 5 gives equivalent data for the respondents to the March 1997 onsite survey of library users.)

7.2.1 Responses to CWeb Online Survey

As the following table shows, for the 80 users of online books, the mean is 14.8 hours spent online per week and the median is ten hours per week online. The greatest number of hours online reported was 71.

Table 55. CWeb Online Survey: Weekly Hours In Online Activities, September 1996 - June 1997

Hours/Week in Online Activities	Number of Respondents	% of Respondents
Less than 2	2	2%
2-4	11	14%
4-6	7	9%
6-8	11	14%
8-10	10	12%
10-12	10	12%
12-14	3	4%
More than 14	26	32%

Breaking down the sample into those who claimed easy access to a networked computer (94%) and those who did not, gives means of 15 and 12.3 weekly hours online, respectively. This is not significant because of the small sample size.

7.2.2 Responses to Questionnaire with Paper *OED*

As the following table shows, for users of the paper format, online activity is lower with a mean of just 3.9 hours.

Table 56. Weekly Hours In Online Activities for 11 Respondents to Paper Questionnaire on Use of OED, 1996 - Percent of Respondents

Hours/Week In Online Activities	% of Respondents
1-2	36%
2-4	36%
4-6	9%
6-8	9%
Extremes (Max. 17)	9%
Mean	3.9

This set of findings on time in online activities by type of resource being used supports a hypothesis that users of online books will be people who spend significantly more hours per week in online activities than do users of the paper versions. As we reported earlier, onsite library users reporting on their use of online resources in the average week in Winter 1997 had a mean of 5.8 hours.

7.2.3 Responses to In-Class Questionnaire, Fall 1996 and Spring 1997

The in-class questionnaire also asked about weekly hours in online activities. (See Exhibit 4.) Responses were distributed as follows.

Table 57. In-Class Surveys: Weekly Hours In Online Activities: Whole Sample Fall 1996 and Spring 1997

Hours/Week	Fall 1996 (N=398)	Spring 1997 (N=217)
1-2	26%	34%
2-4	30%	27%
4-6	19%	13%
6-8	8%	10%
8-10	5%	3%
10-12	4%	3%
12-14	1%	3%
Extremes	8%	6%
Mean	5.2	5.3
Maximum Hours	95	50

Breaking this group down by type of class, we find:

Table 58. In-Class Surveys: Mean Weekly Hours In Online Activities By Class Type: Fall 1996 and Spring 1997

Class Type	Fall 1996		Spring 1997	
	N	Mean	N	Mean
Contemporary Civilization	281	5.1	104	5.8
Political Science	61	4.3	16	6.9
Social Work	56	6.7	97	4.6

The differences among these groups of students and between the two semesters are not statistically significant.

Breaking the Spring 1997 responses down by their reported access to networked computers gives the following table:

Table 59. In-Class Surveys: Weekly Hours In Online Activities By Access to Networked Computer: Spring 1997

	Easy Access to Networked Computer?	
Hours/Week	Yes (N=154)	No (N=55)
1-2	27%	53%
2-4	28%	24%
4-6	16%	7%
6-8	14%	0%
8-10	3%	6%
10-12	3%	6%
12-14	4%	0%
More than 14	7%	6%
Note: Detail may not sum to 100% due to rounding.		

Students claiming easy access to a networked computer spend more time online than those who do not feel that they have easy access. About 31 percent of the former group spend at least six hours a week online while only 18 percent of the latter group do.

In summary, the students surveyed in class have slightly greater uses of online activities than the users of the paper *OED*, surveyed in the library, but about 10 percent less than library users overall in March 1996 and 1997. Students with easy access to a computer are even greater users of online resources. Together these findings suggest that we are at the beginning edge of the transition to electronic use, and makes us confident that we will be able to map out the complete change in attitudes and behavior as availability and accessibility of online books changes the environment at Columbia University.

One of the more interesting issues to be tracked over the span of this project is whether the *center of gravity* of user behavior shifts. That is, will we find in two years that the *light* users of online services are spending more hours per week online than are the average users of today.

* * *

Cover Page **Table of Contents** **Next Page** **Footnotes**

Previous Page

For additional information about the conference, or The Andrew W. Mellon Foundation's scholarly communication initiatives, please contact Richard Ekman. For additional information about ARL or this web site contact Patricia Brennan, ARL Program Officer at (202) 296-2296.

Return to Office of Scholarly Communication Home Page

ARL Home ARL Scholarly Communication and Technology Home Page

© Association of Research Libraries, Washington, DC
Web Design by *Angelo F. Cruz*
Maintained by *ARL Web Administrator*
Last Modified: September 8, 1997

Scholarly Communication and Technology

Conference Organized by The Andrew W. Mellon Foundation

at Emory University
April 24-25, 1997

Copyright © of the papers on this site are held by the individual authors or The Andrew W. Mellon Foundation. Permission is granted to reproduce and distribute copies of these works for nonprofit educational or library purposes, provided that the author, source, and copyright notice are included on each copy. For commercial use, please contact Richard Ekman at the The Andrew W. Mellon Foundation.

Session #4 Patterns of Usage

Online Books at Columbia: Measurement and Early Results on Use, Satisfaction, and Effect

Carol A. Mandel
Deputy University Librarian
Columbia University

and

Mary C. Summerfield
Coordinator, Online Books Project
Columbia University Libraries

and

Paul Kantor
Consultant

……

8. CONCLUSIONS

We find ourselves at the beginning of a complex project, which we have planned in considerable detail. We are not as far along as we would have hoped, principally because it is much more difficult to bring books into online form than everyone has anticipated. In addition, the explosive growth of Web browser technology creates continuous pressure to improve the functionality of the books that go online. We are in the position of surgeons who must keep a patient active, while making drastic changes to that patient's anatomy. We believe that our procedures for surveying the users and analyzing their responses will help us to do that. We hope, in addition, that the results will prove useful to others who wish to understand or to replicate the transitions being pioneered at Columbia.

We look forward to suggestions and comments from others who are approaching this problem from other perspectives.

Appendix 1. Contemporary Books In the Online Collection

Publisher/Title	Author	Subject	Print Status	Month Public	No. of Chapt. or Essays	No. Of Pages	Price - Hard ($)	Price - Soft ($)
Columbia University Press								
Great Paleozoic Crisis	Erwin	Earth Science	Circulati	6/97	10/c	327	57.00	27.50
Seismosaurus: The Earth Shaker	Gillette	Earth Science	Circulati	10/96	11c	205	39.95	
Invasions of the Land	Gordon	Earth Science	Circulati	6/97	10c	312	65.00	17.50
Folding of Viscous Layers*	Johnson	Earth Science	Circulating			461		
Dinosaur Tracks and Other Fossil Footprints*	Lockley	Earth Science	Circulating			338		
Sedimentographica: Photographic Atlas	Ricci-Lu	Earth Science	Circulati	1/97	8c	255	45.00	
Development of Biological Systematics*	Stevens	Earth Science	Circulating			616		
Consuming Subjects*	Kowales	Economic History	Circulating			185	39.50	15.50
Jordan's Inter-Arab Relations	Brand	Internat'l Relations	Reserves	3/97	8c	350	45.00	
Managing Indonesia	Bresnan	Internat'l Relations	Reserves	3/97	10c	375	55.00	19.50
Logic of Anarchy*	Buzan	Internat'l Relations	Reserves			267		

Hemmed In: Responses to Africa's...	Callagy	Internat'l Relations	Reserves	3/97	13e	573	52.50	19.50
China's Road to the Korean War*	Chen	Internat'l Relations	Circulating			339	37.50	17.50
Culture of National Security*	Katzenste	Internat'l Relations	Reserves			562	50.00	17.50
International Relations Theory & the End of the Cold War*	Lebow	Internat'l Relations	Circulating			2925	45.00	17.50
The Cold War on the Periphery*	McMaho	Internat'l Relations	Circulating			431	34.50	17.50
Losing Control: Sovereignty...*	Sassen	Internat'l Relations	Circulating			149	24.95	
Gender In International Relations	Tickner	Internat'l Relations	Reserves Circulating	11/96	5c	180	30.00	15.50
The Inhuman Race*	Cassuto	Literary Criticism	Circulating			289	49.50	17.50
Rethinking Class: Literary Studies...*	Dimock	Literary Criticism	Circulating			285	49.50	16.50
The Blue-Eyed Tarokaja*	Keene	Literary Criticism	Circulating			210		24.50
Ecological Literary Criticism*	Kroeber	Literary Criticism	Circulating			185	49.50	16.50
Parables of Possibility*	Martin	Literary Criticism	Circulating			263	27.50	
The Text and the Voice*	Portelli	Literary Criticism	Circulating			415		
At Emerson's Tomb*	Rowe	Literary Criticism	Circulating			320	49.50	16.50
Extraordinary Bodies: Figuring Physical...*	Thomson	Literary Criticism	Circulating			200	45.00	16.50
What Else But Love? The Ordeal of Race...*	Weinstei	Literary Criticism	Circulating			237	42.00	15.50
Columbia Granger's Index to Poetry	Granger	Poetry	Ref. Desk	10/94	N/A	231		
Ozone Discourses	Liftin	Political Science	Reserves	1/97	6c	257		16.50

Concise Columbia Electronic Encyclopedia		Reference	Ref. Desk	3/91	N/A	943		49.95
Hierarchy Theory*	Ahl	Science	Circulating			206	49.50	
Refiguring Life: Metaphors of 20th Century Biology*	Keller	Science	Circulating			134	20.00	16.50
The Molecular Biology of Gaia*	Williams	Science	Circulating			210	45.00	
Sampling the Green World*	Stuessy	Science/ Botany	Circulating			289	49.50	
The Illusion of Love*	Celani	Social Work	Circulating			217	27.50	16.50
Mutual Aid Groups, Vulnerable Populations, & the Life Cycle	Gitterma	Social Work	Reserves	11/96	21e	448	39.50	
Supervision in Social Work	Kadushin	Social Work	Reserves	9/96	10c	597	45.00	
Eating Disorders: New Directions*	Kinoy	Social Work	Circulating			166	42.50	15.00
From Father's Property to Children's Rights*	Mason	Social Work	Circulating			237	40.00	16.00
Handbook of Gerontological Services	Monk	Social Work	Reserves	9/96	25e	694	65.00	
Turning Promises Into Performance	Nathan	Social Work	Circulati	9/96	13c	160	45.00	15.50
Philosophical Foundations of Social Work	Reamer	Social Work	Reserves Circulati	9/96	5c	219	49.50	17.00
Task Strategies: An Empirical Approach	Reid	Social Work	Reserves	9/96	11c	329	37.50	
Experiencing HIV*	Sears	Social Work	Circulating			182	45.00	15.00
Qualitative Research In Social Work	Sherman	Social Work	Reserves	1/97	43e	520	57.50	27.50
The Empowerment Tradition in America*	Simon	Social Work	Circulating			297	49.00	22.50

Garland Publishing

Native American Women	Bataille	Biograph	Referenc	1/97	N/A	333		
African American Women	Salem	Biograph	Referenc	1/97	N/A	622		
Chaucer Name Dictionary	de Weever	English Literature	Referenc	12/96	N/A	451		
Oxford University Press								
Oxford English Dictionary		Language	Referenc	9/96	N/A	N/A		
Postcards from the Trenches: Negotiating the Space Between Modernism & The First World War**	Booth	Literary Criticism	Circulating			186	35.00	
The Erotics of Talk**	Kaplan	Literary Criticism	Circulating			240?	35.00	16.95
"Littery Man": Mark Twain...	Lowrey	Literary Criticism	Circulati	11/96	4c	177	39.95	
Children's Literature & Critical Theory	May	Literary Criticism	Circulati	11/96	9c	243	29.95	18.95
Poetics of Fascism	Morrison	Literary Criticism	Circulati	10/96	4c	177	39.95	
Novel & Globalization of Culture	Moses	Literary Criticism	Circulati	11/96	4c	240		18.95
Modernism & the Theater of Censorship	Parkes	Literary Criticism	Circulati	6/97	4c	242	45.00	
Romances of the Republic: Women, the Family, & Violence in the Literature of the Early American Nation**	Samuels	Literary Criticism	Circulating			208	39.95	
Majestic Indolence: English Romantic Poetry...	Spiegelm	Literary Criticism	Circulati	1/97	6c	221	49.95	
Making Mortal Choices: Three Exercises in Moral Casuistry**	Bedau	Philosoph	Circulating			123	29.95	13.95

Morality, Normativity & Society	Copp	Philosoph	Circulati	10/97	10c	262	42.00	
Free Public Reason: Making It Up ...	D'Agosti	Philosoph	Circulati	10/96	10c	203	45.00	
Metaphilosophy and Free Will**	Double	Philosoph	Circulating			192?	35.00	
Bangs, Crunches, Whimpers & Shrieks	Earman	Philosoph	Circulati	10/96	8c	257	35.00	
Causation and Persistence: A Theory of Causation**	Ehring	Philosoph	Circulating			191	39.95	
Self Expression: Mind, Morals & Meaning...	Flanagan	Philosoph	Circulati	10/96	12c	222	24.95	
Logic of Reliable Inquiry	Kelly	Philosoph	Circulati	10/96	16c	434	59.00	
Philosophy of Mathematics & Mathematical Practice In the 17th Century	Mancosu	Philosoph	Circulati	10/96	6c	275	60.00	
Moral Dilemmas & Moral Theory**	mason	Philosoph	Circulating			246	45.00	
Autonomous Agents: From Self Control to Autonomy	Mele	Philosoph	Circulati	10/96	13c	371	49.95	
Other Minds: Critical Essays	Nagel	Philosoph	Circulati	10/96	22e	229	26.00	
The Last Word**	Nagel	Philosoph	Circulating			147	19.95	
Law & Truth	Patterson	Philosoph	Circulati	11/96	7c	189	39.95	
Nietzsche's System	Richards	Philosoph	Circulati	10/96	4c	316	35.00	
Freedom & Moral Sentiment	Russel	Philosoph	Circulati	10/96	12c	200	45.00	
Living High and Letting Die**	Unger	Philosoph	Circulating			187	39.95	14.95
The Human Animal**	Weston	Philosoph	not in yet			208?	29.95	
Real Rights	Wellman	Philosoph	Circulati	10/96	8c	279	52.00	

Simon & Schuster

Bond Markets**#	Fabozzi	Business	Reserves			595	72.00	
Marketing Management**#	Kotler	Business	Reserves			824	75.00	
Statistics for Business & Economics**	Newbold	Business	Reserves			895	61.50	
Investments**	Sharpe	Business	Reserves			900	80.00	
Financial Market Rates & Flows**	Van Horne	Business	Reserves			338	41.00	
Politics & the Media	Davis	Political Science	Reserves	4/97	9c	432	30.00	
Public Policy Analysis**	Dunn	Political Science	Reserves			480		
International Politics	Holsti	Political Science	Reserves	1/97	15c	432	32.25	

Notes: * Permission has been received, but the book is not yet online. ** Book is not yet online.

# A new edition has been issued for which we need the electronic file.

Appendix 2. Format Availability, Winter 1997, and Introduction Date

	Unix w/Patty	**CNet - Own format/ Patty**	**CNet w/Lynx text-based browser**	**CWeb w/Lynx text-based browser**	**CWeb w/ graphical browser**
Columbia Concise Electronic Encyclopedia.		3/91			
Oxford English Dictionary	8/94	8/94		9/96	9/96
Granger's Index to Poetry			10/94	10/94	10/94
Chaucer Name Dictionary			12/96	12/96	12/96
Native American Women			2/97	1/97	1/97
African American Women			2/97	2/97	2/97
All other books				1995-1997	1995-1997

- *Unix access* allows the user to simple type in a command (e.g., *OED*) at the Unix prompt and to go to the resource from there. The Patty 2.0 interface and search engine allows sophisticated searching. In the Unix mode the user can save a whole entry to a file.

- *ColumbiaNet (CNet)* is a gopher-based Campus Wide Information System (CWIS). Users can *clip* individual screens (which may not encompass a whole entry) and download (file transfer), mail or print them. Some Web documents have been linked to CNet menus via *lynx*, the text-based Web browser. The user then has both lynx and CNet functionality.
- *ColumbiaWeb (CWeb)* is Columbia's main World Wide Web site. LibraryWeb (LWeb) is the Libraries' site within CWeb. Both can be accessed by either a graphical browser or lynx. All graphical browser functionalities are available to the community while using the online books - reading, finding words, printing pages or whole documents, saving to a file, copying and pasting to a word processor document. A lynx user can save an entire Web file to a local file, email it or print it to the screen. Or she can copy and paste portions of a Web file to another document.

Appendix 3. Home Computers: Minimum Specifications Recommended

Specificatio	5/94 (for student)	12/94	4/95	8/95	11/95
CPU	486	486/66	486DX2/66	75 Mhz Pentium	75 Mhz Pentium (slowest Pentium available)
RAM	4 MB	8 MB	8 MB	8 MB	8 MB
Hard Drive Capacity	100 MB	340 MB	500 MB	750 MB	1 GB
CD-ROM Drive		Double	Double	Quad	Quad
Monitor: Size Capacity	SVGA-compatible		15" 256-color SVGA	17" better	15" 72-Hz
Graphics: Pixels Colors Video RAM	SVGA		640 x 480 256	65,000 1 MB	640 x 480 256 1 MB
Sound: Card Speakers			16-bit Sound Blaster-compatible	Sound Blaster comp. Stereo	16-bit SB-compatible Powered
Modem			14.4 kbps fax modem	14.4 kbps	14.4 kbps fax modem
Price	$1,500	$2,500	$1,800-$2,00	Not Given	About $2,000
Source	**Family PC,** 5/1/94, p. 140+	**PC Magazine**	"Family Shopper Smartcard," **Family PC,** 4/95	Walter S. Mossberg, *Personal Technology*, **Wall Street Journal**, 8/31/95, p. B1	"Multimedia Family PCs: New Minimum System Requirements," & "Family Shopper Smartcard," **Family PC,** 11/17/95

Specification	4/96	4/96	5/96	11/96	4/97
CPU	75 Mhz Pentium	133 Mhz Pentium	120 Mhz Pentium	133 Mhz Pentium (200 Mhz is fastest available)	166 Mhz MMX Pentium (200 Mhz if possible)
RAM	8 MB	16 MB	16 MB	16 MB EDO RAM	16 MB (32 preferable)
Hard Drive Capacity	1 GB	Fast 1.2 GB	1.2 GB	1.2 GB	2.0 GB, more if possible
CD-ROM Drive	Quad	Quad	Quad	6X (up to 12X available)	8X
Monitor: Size	15" 72-Hz	15" (17" better)	17"	15"	17" with maximum .28 mm dot pitch
Graphics: Pixels Colors Video RAM	640 x 480 256 1 MB	64 bit Graphics Accelerator	65,000 1 MB	Accelerated PCI graphics card w/1 MB RAM	2 MB VRAM
Sound: Card Speakers	16-bit SB-compatibl Powered	Not mentioned	16-bit SB-compatibl	16-bit FM music synthesis, SB-compatible	
Modem	14.4 kbps fax modem	28.8 kbps data/fax modem	28.8 kbps	28.8 kbps	33.5 kbps, upgrade to 56 kbps
Connectors					Pair USB ports
Price	About $2,000	$2,500 or more	$2,000-$2,5	$2,000	Not given
Source	"Family Shopper Smartcard," **Family PC**, April 1996	Bill Howard, "At Home," **PC Magazine,** 4/23/96, p. 300. Min. For "Perfect Home Computer"	Walter S. Mossberg, *Personal Technology*, **Wall Street Journal**, 5/10/96, p. B1	*"The '97 Multimedia Family PC,"* **Family PC**, Dec. 1996 (recd 11/19/96), p. 68.	Walter S. Mossberg, *Personal Technology*, **Wall Street Journal**, 4/9/97, p. B1

Appendix 4. Gateway 2000 Home Computers: System Specifications

Specification	12/94	4/95	5/95	12/95	4/96
Computer Name	P5-60 Family PC (featured in ads)	P5-60 Family PC (featured in ads)	P5-75 Family PC (featured in ads)	P5-100 Family PC (Top Rated, Best Buy)	P5-150
CPU	60 Mhz Pentium	60 Mhz Pentium	75 Mhz Pentium	100 Mhz Pentium	150 Mhz Pentium
RAM	8 MB	8 MB	8 MB	8 MB	16 MB
Hard Drive Capacity	540 MB	540 MB	730 MB	1 GB	1.6 GB
CD-ROM Drive	2x	4x	4x	4x	6x
Monitor: Size	14"	15"	17"	15"	17"
Graphics: Pixels Colors Video RAM	1 MB	1 MB	2 MB	2 MB	2 MB
Sound: Card	16-bit SB-compatibl	16-bit SB-compatibl	16-bit SB-compatibl	16-bit SB-compatibl	16-bit SB-compatible
Speakers	Altec Lansing	Altec Lansing	Altec Lansing	Altec Lansing ACS-40	Altec Lansing
Modem	14.4 kbps fax modem	14.4 kbps fax modem	14.4 kbps fax modem	28.8 kbps fax modem	28.8 kbps fax modem
Software Included	MS Works + CD-ROMs	MS Works + CD-ROMs	MS Works + CD-ROMs	MS Works, 20 CD-ROMs	MS Office 95
Price (+ shipping)	$2,099	$2,099	$2,499	$2,149	$2,899
Source	GW2000 ad	GW2000 ad	GW2000 ad	"Multimedia Family PCs: Recommended Systems," **Family PC**, 11/17/95 & GW2000 advertising insert, 12/95	GW2000 ad

Specification	**5/96**	**5/96**	**10/96**	**10/96**	**5/97 & 6/97**	**5/97**
Computer Name	P5-133 Family PC	P5-120 Family PC	P5-166 Family PC (featured in ads)	G6-180 Family PC (featured in ads)	G5-166M (featured in ads)	G5-200M (modified)
CPU	133 Mhz Pentium	120 Mhz Pentium	166 Mhz Pentium	180 Mhz Pentium Pro	166 Mhz MMX Pentium	200 Mhz MMX Pentium
RAM	16 MB	16 MB	16 MB/256 KB Cache	32 MB/256 KB Internal Cache	16 MB/256 KB Internal Cache	16 MB/512 KB Internal Cache
Hard Drive Capacity	1.62 GB	850 MB	2 GB	2 GB	1.6 GB	1.6 GB
CD-ROM Drive	8x	8x	8x	8x	12X	12X
Monitor: Size	17" Vivitron	15" CrystalScan	17" CrystalScan	17" Vivitron	15" CrystalScan	15" CrystalScan
Graphics Video RAM	64 Bit PCI Accelerator w/MPEG Video Scaler 2 MB	64 Bit PCI Accelerator w/MPEG Video Scaler 1 MB	64 Bit PCI Graphics 2 MB	64 Bit PCI Graphics 2 MB	64 Bit PCI Graphics 3-D 2 MB	64 Bit PCI Graphics 3-D 2 MB
Sound: Card Speakers	16-bit SB-compati Altec Lansing ACS-40	16-bit SB-compati Altec Lansing ACS-40	16-bit Gateway Sound, SB-comp. Altec Lansing ACS-40	16-bit Ensoniq Wavetable Altec Lansing ACS-40	16-bit Gateway Sound, SB comp. Altec Lansing ACS-41	16-bit Gateway Sound, SB comp. Altec Lansing ACS-41
Modem	28.8 kbps fax modem	28.8 kbps fax modem	36.3 kbps fax modem w/speakerp	36.3 kbps fax modem w/speakerp	Telepath modem w/speakerp x2 Tech upgradeable	Telepath modem w/speakerphone, x2 Tech upgradeable
Software Included	MS Works + 14 CD-ROMs	MS Works + 14 CD-ROMs	Generations II	Generations II	MS Office 97, Encarta, MMX Tech-Enhanced Software	MS Office 97

Price (+ shipping)	$2,549	$1,999	$2,199	$2,649	5/97: $1,799 6/97: $1,749	$2,064
Source	GW2000 ad, 5/96	GW2000 ad, 5/96	GW2000 ad, 10/96	GW2000 ad, 10/96	GW2000 ads	GW2000 website

Appendix 5. U.S. Household PC & Internet Penetration

(ALL PERCENTAGES)

Year & Source *(so)* for Estimate	*2000*	*so*	*199*	*so*	*Jan -97*	*so*	*Dec -96*	*so*	*Sep -96*	*so*	*Jan -96*	*so*	*Sep -95*	*so*	*Jun -95*	*so*	*Jan -95*	*so*	*Sep -94*	*so*	*Mar -94*	*so*
HH Penetrati PCs	55	8	47	8	37	1	39	8	37	9	35	1	35	9	32	1	31	1	35	9	31	3
					35.	6					34	2			36	3	33	2				
											32.	6										
First Time Buyers of Home PCs Share of Total Sales Home PCs					32	1					50	1									50	1
Share of Primary HH PCs:																						
Pentium					40%	10					24	2			7	3	4	2				
486					60%	10									40	3						
Windows-based PCs (Share of All HH PCs)					59.	6					42.	6										
HH Penetrati Modems	50	8	42	8			28	8			>5	6			18	1	16	1			12	3
											70	2			20	3	73	2*				
																	44.	6				

HH Penetrati Online Services	35	8	27	8	17	1	17	4	14	5	11	1,4	6.4	9	9	1			4.5	9	5	3
							15	8	10.	9	46	2**			12	3	25	2*				
HH Penetration of Other Internet Access											26	2**										
Adults (>=age 16) using Internet							~25	7			<1	7										
HH Penetration CD-ROMs											70	2			15	3						
											83	2*					55	2*				

NOTES:

* Percentage of new computers purchased in previous year.

** Percentage of modem owners subscribing to an online service.

SOURCES:

1. Odyssey Ventures. Jared Sandberg, "PC Makers' Push Into More Homes May Be Faltering," Wall Street Journal, March 6, 1997
2. Software Publishers Association Surveys (Telephone survey, about 500 adults)
3. Times Mirror Surveys (Telephone survey, about 3,600 adults)
4. Electronic Information Report, *1996 Online Subscriber Survey*
5. PC-Meter, Use of Internet Access Service in Last Month, reported by Reuters in clari-news
6. NPD Group, Inc., http://www.npd.com/meterpr4.htm, "Latest NPD Survey Finds World Wide Web Access From Homes Grew Fourfold in Second Half of 1995" Report on survey of sample of 44,800 homes.
7. Nielsen Media Research survey for CommerceNet, Julia Angwin, "Internet Usage Doubles in a Year, San Francisco Chronicle, March 13, 1997
8. Jupiter Communications, http://www.jup.com/Jupiter/release/jan97/consumer.shtml, January 6, 1997 "New Devices and Technologies Will Drive Net Into 36 Million Homes by 2000"
9. INTECO Corp., http://www.inteco.com/pu961031.html, "Percent of HH with PCs, Internet Access: 1993 to 1996

10. PC-Meter, "Pentium, Windows 95 in Minority of Home Installs, According to New PC Meter Reports," July 1, 1997 Press Release from PC-Meter. PC Meter panel consists of 10,000 PC-owning households in the U.S. Note: Value given as share of 486 processors is actually all sub-Pentium PC processors, so includes 286s and 386s as well. Report also notes that 47 percent of household PCs now have Windows 95.

Appendix 6. Concise Columbia Electronic Encyclopedia Sessions, 1994 - 1997: CNet

	Number of Sessions				% Change:		
Month	**1994**	**1995**	**1996**	**1997**	**'94 to '95**	**'95 to '96**	**'96 to '97**
Jan.	1,186	993	645	224	-16%	-35%	-77%
Feb.	1,645	1,342	991	341	-18%	-26%	-66%
Mar.	1,529	1,508	997	394	-1%	-34%	-60%
April	1,551	1,335	896	250	-14%	-33%	-72%
May	1,042	757	458	197	-27%	-39%	-57%
June	713	466	326		-35%	-30%	
July	712	475	412		-33%	-13%	
Aug.	708	496	251		-30%	-49%	
Sept.	1,160	761	291		-34%	-62%	
Oct.	1,879	1,251	542		-33%	-57%	
Nov.	1,596	1,029	405		-36%	-61%	
Dec.	1,300	833	300		-36%	-64%	
Total	15,021	13,2416	6,514		-25%	-42%	

Note: July 1995 hits are estimated.

* * *

Cover Page Table of Contents Next Page Footnotes

Previous Page

For additional information about the conference, or The Andrew W. Mellon Foundation's scholarly communication initiatives, please contact Richard Ekman. For additional information about ARL or this web site contact Patricia Brennan, ARL Program Officer at (202) 296-2296.

Return to Office of Scholarly Communication Home Page

ARL Home ARL Scholarly Communication and Technology Home Page

© Association of Research Libraries, Washington, DC

Web Design by Angelo F. Cruz
Maintained by ARL Web Administrator
Last Modified: September 8, 1997

Scholarly Communication and Technology

Conference Organized by The Andrew W. Mellon Foundation

at Emory University
April 24-25, 1997

Copyright © of the papers on this site are held by the individual authors or The Andrew W. Mellon Foundation. Permission is granted to reproduce and distribute copies of these works for nonprofit educational or library purposes, provided that the author, source, and copyright notice are included on each copy. For commercial use, please contact **Richard Ekman** at the The Andrew W. Mellon Foundation.

Session #4 Patterns of Usage

Online Books at Columbia: Measurement and Early Results on Use, Satisfaction, and Effect

Carol A. Mandel
Deputy University Librarian
Columbia University

and

Mary C. Summerfield
Coordinator, Online Books Project
Columbia University Libraries

and

Paul Kantor
Consultant

• •

FOOTNOTES:

[1] The book could be used entirely in an online format or the scholar could choose to acquire a print version of all or part of the book once he had browsed the online version. Alternatively, at least at some point in time and for some forms of books, such as textbooks, an electronic format such as a CD-ROM might be better - for technical, cost or market reasons - than either the online or the print format. Malcolm Getz addressed some of the format issues well in his paper at this conference, *Electronic Publishing in Academia: An Economic Perspective.*

[2] In effect, funds that would have been spent on interlibrary loan activities, i.e., staff and mailing costs, would be redirected to the producers of the scholarly knowledge, thus supporting the production and dissemination of such scholarship.

[3] Detailed background information is provided in the Project's *Analytical Principles and Design* document of December 1995 and in its Annual Report of February 1997. Both are available at http://www.columbia.edu/dlc/olb/.

[4] Ultimately, if online books were to become a regular product of scholarly publishers, the publishers would make the online version a regular output of their production process. This reengineering might or might not lead to a reduction in publishers' production costs, but it would certainly mean that universities would not be faced with the conversion of printers' tapes to HTML.

[5] Software allowing annotation of an electronic document is available, but few people are aware of it. The Project will seek to bring such software to the Columbia community as feasible.

[6] A few reference books were already online. Their design will be discussed shortly.

[7] The SGML mark-up of these texts as provided was inconsistent. A conversion which was expected to be done quickly and nearly automatically was instead a labor intensive, time consuming process, resulting in the delayed provision of the texts on the Web.

[8] Publishers, including Chadwyck-Healey and Oxford, have provided permission for such conversion.

[9] Greater detail is available in the Annual Report. See Appendix 1 for a summary of the online collection with titles, subject matter and location within the Libraries collections of the print copies (reference, regular circulating collection, reserves collection).

[10] Until recently contracts with authors contained no provision for electronic versions of books. Current contracts include such provisions, but royalties are specified as a percentage of revenues derived from sales in electronic format. As there is no price for the materials included in this research effort, the Press needs to obtain permissions for this special use.

[11] Our agreement with Garland requires them to provide HTML-coded files as we do not have funding to undertake conversion for additional publishers. Thus, Garland must assess its manpower and funding availability in determining the books it will provide.

[12] One Oxford file is awaiting conversion.

[13] Two of these books, designed for course use, have gone into new editions since we received the electronic files, so we need to obtain the files for the latest editions before putting these two books online.

[14] See the Annual Report on the Project's Web home page for greater detail.

[15] So far the Project Coordinator has conducted telephone interviews with three authors who refused permission and inquired about their reasons for doing so. Columbia University Press received explanatory comments from several other authors when they refused permission.

[16] Recommendations from satisfied users to their colleagues is always one of the key sources of sampling for new products. However, it is one over which we have least control.

[17] By early Summer 1997, the online Reserves catalog will contain entries for the online versions of books that have been put on reserve for various courses. This should increase usage by students in courses for which a book is assigned reading.

[18] A repeat user of the collection could bookmark any of the pages and return to it with just one step.

[19] *The OED, Granger's Index to Poetry, Chaucer Name Dictionary, African American Women*, and *Native American Women*.

[20] Consistent with marketing theory, the TULIP project found that usage of online journals was much greater at institutions that had conducted substantial campaigns to build awareness and trial. See Elsevier Science, *TULIP Final Report*, 1996, for details.

[21] As noted earlier, this document is included on the Web page for the Project. Questionnaires and other research methodologies have been fine-tuned after pretests and early use, but the general concepts remain in place.

[22] Networked printing that allows such tracking is not yet in place.

[23] The chart does not have value labels, so these are estimates of the values.

[24] Source: Amy Cortese, "A Census in Cyberspace," *Business Week Online News Flash*, April 24, 1997

[25] Data are taken from a fact sheet issued by Columbia's Academic Information Systems in February 1997.

[26] A question about access to a Web browser has just been added to Project questionnaires.

[27] In a Spring 1997 interview, a second year social work graduate student who lives in New Jersey said that she has a modem in her home computer but does not use it as direct dialup to Columbia is a long distance call and at $20 a month an ISP account is too expensive. She thought that about half her classmates might not use the Internet from home even if they had computers.

[28] For the following tables, Web data prior to May 1966 include hits by Project staff, those from May 1966 forward do not. These were excluded as they can be substantial in number as resources are in design phases and do not reflect the scholarly use that we are studying. The earlier statistics cannot be refined to extract such hits. NA: data are not available; NC: total or change is not calculable.

[29] Analyzing server data is proving more difficult than anticipated as user identification is instituted. In the future, data will be reported on a monthly basis and reports will be issued quarterly.

[30] This is the common opinion of users who have completed questionnaires and others who were interviewed by email and in person.

[31] While AcIS has designed a Web version of *The OED* which has various analytical capabilities, unfortunately that version requires more server resources than AcIS can devote to

this single work.

[32] These are host computers with addresses linking them to the dormitory network. See section 4.3.1.2.

[33] Their questionnaires have not been removed since we started tracking these books in January 1997.

[34] A larger collection is available on CD-ROM in the Electronic Text Service. We could not obtain permission to put some of these online.

[35] This type of skewed distribution, or Bradford law, is typical of all types of library collections.

[36] Gitterman's *Mutual Aid Groups* was on reserve for three professors for Fall 1996 semester, but we could not put it in the collection accessible to the Columbia community until November due to delays in obtaining permission from the authors. It was in use in three Social Work classes for Spring 1997 semester.

[37] Community members can request that a book be recalled but that process is not guaranteed to bring the book back to the library and it can take several weeks. Online books are always accessible for such browsing.

[38] This impression was confirmed in a recent interview with a second year social work student. She noted that her home computer has a modem but that she does not have a ISP account or dial-in to Columbia from her home in New Jersey (a long distance call) because of the cost. All of her use of online resources occurs on campus, in the Social Work computer lab or in a library.

[39] Hitherto, IP addresses were the basis of controlling access to the books. The new authorization system was put into effect for *The OED* at the first of April 1997. One means of accessing *The OED*, via a bookmark, is necessarily not included, so *The OED*'s use is still understated. It is not yet in place for *Granger's Index to Poetry* so that resource is not included in this analysis.

[40] We are in the process of obtaining permission from institutions affiliated with Columbia to use the directory information about their users for our research. As a result, we do not have cohort detail on 40 (14%) of the users. Percentages given are with that set of 'unidentified users' as a separate group.

[41] We have removed Officers of the Libraries and specific other individuals who are involved in the Project from these data.

[42] For use data to show revealed preference, the collection must contain books that would draw users to the collection repeatedly - either books that users want to look at repeatedly or an assortment of books that pulls scholars to the collection for a variety of purposes.

[43] The *Past Masters* texts used in the second semester of the course are studied at the beginning of the semester making it difficult to make arrangements with the instructors on time.

[44] JAKE is the networked laser printer system maintained by AcIS. Undergraduates and Social Work students have a free 100 page quota for JAKE printing each week.

[45] The two Garland reference books are not separated out, even though different uses are offered on their questionnaires.

[46] Two changes were instituted in the middle of the Spring semester - a snappier line requesting completion of the survey and a change to a $20 gift certificate from a $20 copycard.

[47] All textual documents available on CWeb can be accessed via any graphical interface or via *lynx*. Lynx can be used at the University's Unix prompt with the command *lynx http://www.columbia.edu/....../*

• •

Cover Page   Table of Contents   Page 1 of Report

Previous Page

For additional information about the conference, or The Andrew W. Mellon Foundation's scholarly communication initiatives, please contact Richard Ekman. For additional information about ARL or this web site contact Patricia Brennan, ARL Program Officer at (202) 296-2296.

Return to Office of Scholarly Communication Home Page

ARL Home ARL Scholarly Communication and Technology Home Page

© Association of Research Libraries, Washington, DC
Web Design by Angelo F. Cruz
Maintained by ARL Web Administrator
Last Modified: September 8, 1997

Scholarly Communication and Technology

Conference Organized by The Andrew W. Mellon Foundation

at Emory University
April 24-25, 1997

Copyright © of the papers on this site are held by the individual authors or The Andrew W. Mellon Foundation. Permission is granted to reproduce and distribute copies of these works for nonprofit educational or library purposes, provided that the author, source, and copyright notice are included on each copy. For commercial use, please contact **Richard Ekman** at the The Andrew W. Mellon Foundation.

Cocktails and Dinner, Michael C. Carlos Museum

Digital Documents and the Future of the Academic Community

Peter Lyman
Professor, School of Information Management Systems, and University Librarian
The University of California, Berkeley

……

Digital Documents and the Future of the Academic Community [*]

Today the academic community is the subject of an experiment in technological innovation. That experiment is the introduction of digital documents as a new currency for scholarly communication, an innovation which will perhaps replace, or perhaps complement the system of print which has evolved over the past century. What are the long term consequences of this innovation for the conduct of research and teaching, for the library and the campus as organizations and places, and ultimately for our sense of academic community?

This conference on Scholarly Communication and Technology has primarily focused upon one key dimension of this process of innovation, the economics of scholarly publishing. The central focus has been on the issue of the cost and availability of information: Will digital modes of publication be more cost effective than print, both for publishers and for libraries? But other questions are implicit as well. We has discussed how readers use on-line journals, and how the journal format itself is evolving in a digital medium: Will digital publications change the form and content of scholarly ideas? Together, these papers investigate the emerging outline of a new marketplace for ideas, one which perhaps will yet be reshaped by new kinds of intellectual property law, but which certainly will include new kinds of pricing, new products, and new ways of using information. These are certainly important economic questions, but if we knew the answers would we know enough to understand the dynamics of change in scholarly publishing, and the impact of technological innovation upon the academic community for which the system of scholarly communication serves as an infrastructure?

One reason this question must be asked is the debate about what economists call "the productivity paradox." This is the observation that the introduction of information technology into the office has not increased the productivity of knowledge workers thus far, unlike the productivity gains which technology has brought to the process of industrial production. Yet Peter Drucker has described the productivity of knowledge workers as the key management problem of the 21st Century.[1] And more recently Walter Wriston has described information as a new kind of capital which will be the key to wealth in the economy of the future, saying: "The pursuit of wealth is now largely the pursuit of information, and the application of information to the means of production."[2] Why, then, has information technology not increased the productivity of knowledge workers? Does it not bring about cultural and organizational changes?

Erik Brynjolfsson has defined three key dimensions within which an explanation for the paradox might be found.[3] First, perhaps this is a problem of measurement, since the outcomes of work mediated by information technology may not fit traditional categories, and are perhaps difficult to measure with traditional methods. Secondly, the productivity paradox might be a consequence of the introduction of very different incentive structures which change the cultures of work, and may require redesign and reorganization of work processes previously based on printed records in order to create productivity gains. And thirdly, perhaps information technology creates new kinds of economic value (such as variety, timeliness and customized service), which change the very nature of the enterprise by introducing new dimensions and qualities of service.

The analysis of the impact of information technology upon scholarly communication has only indirectly been a discussion about productivity thus far, although such a discussion inevitably will begin when it is understood that this will be a discussion about the quality of academic information and work, not just about its efficiency.[4] For the purposes of this discussion, however, what is of immediate interest is the way the productivity issue *frames the possible dimensions of the dynamics of technological innovation*, thereby setting a research agenda for the future. That is: even if digital documents were shown to be more cost effective than printed journals, where might we look to find the consequences of this innovation for the academic community? How might our understanding of the *outcomes* or *impact* of research, teaching and learning change, if at all? How might the *incentives* for academic work evolve, and would the *organization* of the process of research and teaching change? Will new kinds of *value* be introduced into academic work, changing its cultures, and will traditional kinds of value be lost? This is the broader research agenda which provide context for discussion of the price, supply and demand for digital publications.

In sum, how might the substance and organization of academic work change as information technology changes the infrastructure of scholarly communication? To borrow a term from a very different economic tradition, the question of the social impact of information technology concerns the *mode of production*, that is, the complex of social relationships within which academic work is organized, within which the products of academic work are created and consumed, and the cultural valuation given to academic work and its products. In the course of this exploration of the changing modes of production which govern knowledge work it will be necessary to think seriously about whether printed knowledge and digital information are used in the same way, if we are to understand the nature of demand; about the new economic roles of knowledge, if we are to understand issues of price and supply; and about how the management of knowledge might be a strategy for increasing the productivity of knowledge workers.

The system of scholarly communication

The idea that there is a system of scholarly communication was popularized by the ACLS newsletter **Scholarly Communication**, which published a survey on the impact of personal computers upon humanities research in 1985. It is a term invented to frame both print publication and digital communication within a single functional perspective, tacitly asserting a continuity between them. It is this continuity which is in question, not least because the term "scholarly communication" encompasses the very research processes which are being transformed by information technology, creating new kinds of information products and services which were not previously part of the scholarly publishing marketplace. These include, for example, patents on methodological procedures and genetic information; software for gathering, visualizing and analyzing data; information services, such as document delivery and databases; network services; Lists and Web pages; electronic journals and CD-ROM.

Today each of the parts of the system of scholarly communications built over the past fifty years are changing, and it is unlikely that a new equilibrium will resemble the old. This system is unusual, perhaps, in that different participants perceive it from very different, perhaps contradictory, perspectives. From the perspective of the academic community, both the production and consumption of scholarly information are part of a culture of gift exchange. In gift cultures, information is exchanged primarily (although not necessarily exclusively) in order to create and sustain a sense of community greater than the fragmenting force of specialization and self interest. From the perspective of academic publishing, the academic community consists of two markets in which exchanges are governed by contract, that of authors and that of the consumers, the largest of which are academic research libraries. It is this perspectivism, perhaps, which leads each side to hope that digital documents will replace printed journals without changing other aspects of the system of scholarly communication.

Gift and market exchange are symbiotic, not opposites. If scholarly publishing is governed by the rules of market exchange, it must manage the boundaries between two gift cultures, that within which knowledge is created, and that within which knowledge is consumed. The crisis of scholarly communication has made these boundaries very difficult to manage, as ideas from the University are turned into intellectual property, then sold back to the University to be used as a common good in the library.

Why the crisis in boundary management? The immediate crisis which has destabilized the system is the problem of sharply increasing costs for scholarly information. The causes of the crisis are varied, but begin with the commercialization of scholarly publishing, the dramatic shift from nonprofit to nonprofit publishing since the 1950's, creating the hybrid gift/market system described above. In turn, the historic growth in the amount of scientific, technical and medical information, driven by federal funding, has increased costs. And the waning of a sense of the legitimacy of library collection costs within the University has allowed the rate of growth of collection budgets to fall far below the rate of price increases.[5] Even with cost/price increases, the gift economy still subsidizes the market economy, and remarkably those who subsidize research do not yet make an intellectual property claim for the copyrighted intellectual property they support. Subsidies include, for example, the federal funding of research, institutional subsidies, and the voluntary labor of faculty providing editorial services to publishers.

This system evolved at the turn of the 20th Century as a subsidy for non-profit University presses and disciplinary society publishers, in order to circulate scholarly information and build a national intellectual infrastructure. Since 1950, however, federal research funding and commercial publishing have reshaped the system, creating a hybrid market-gift exchange system with many unrecognized cross subsidies.

Higher education is both the producer and consumer of scholarly publications. As creators of scholarship, faculty are motivated by non-market incentives, primarily promotion and tenure; yet at the same time, faculty see themselves as independent entrepreneurs, managing a professional career in self governed disciplines and educational institutions. This guild-like structure is a deliberate anachronism, perhaps, but one which sustains a sense of professional identity through moral as well as material rewards.

Scholarly publications are consumed within a gift culture institution called the library, a subsidized public good within which knowledge appears to the reader as a free good. This gift culture is, in turn, subsidized by the owners of intellectual property through the Fair Use and First Sale doctrines, which generally allow copyrighted information to be consumed for educational purposes.

The ambiguity at the boundary of gift and market extends to institutions of higher education as well, which are simultaneously corporation and community. But the dominant factor which has shaped the last fifty years is that Universities have become a kind of public interest corporation serving national policy goals. Modern research Universities have been shaped by federal research funding since the Sputnik crisis, as "milieux of innovation" to function as tacit national laboratories for a polity uncomfortable with the idea of a formal industrial policy.[6]

This system of scholarly communication is in an irreversible process of change. Consider, for example, the possible consequences for this system if some of the ideas and questions being debated nationally were to come to pass:

- What is the future of University research? Do research Universities still play a central role as national milieux for innovation, or has the corporation become the focus of innovative research and national information policy?
- What is the future scope of higher education? Historically Colleges and Universities have had a tacit monopoly of the education market, based upon accreditation and geographical proximity, but instructional technology and distance education have created new markets

for education. With the Western Governor's University proposal a national market for education would be created, based on selling teaching services and evaluated by examination, which in principle could compete with the traditional institutional settings for education.

- What is the future of the Library as a public good? In the polity, the idea of a national digital library has been modeled upon the universal access policies governing telephone and electric utilities. Here the public good is fulfilled by the provision of "access," but it will be the consumer's responsibility to pay for information used.

- What is the future of Fair Use? Rights which exist in print are not being automatically extended to the use of digital works. Federal policy discussions about intellectual property in the digital environment have not included Fair Use, giving priority to the creation of a robust market in digital publication and the creation of incentives for the publication of educational works.

These are questions, not predictions, but they are questions which are being discussed in the polity, so they are not mere speculation. They are intended only to point out that the system of scholarly communication is a historical creation which was a response to certain conditions which may longer exist.

Three new factors define the conditions within which a system of scholarly communication may evolve. First is the emergence of a global economy in which intellectual property is an important source of wealth, thus the value of scholarly research may be a matter of economic interest extending far beyond the traditional concerns of the academy. Secondly, the end of the cold war as a stimulus for national information policy which took the form of federal patronage of University research may fundamentally change the shape and content of federal funding for research. And thirdly, the astonishing cultural diversity of our society, and the replacement of a melting pot ideal by a transnational culture (in which family, ethnic, corporate and professional loyalties may cross and transcend national boundaries), may create entirely new social contexts for education. For example, outside of the sciences, scholarly disciplines have tended to have national scope, but are likely to develop international paradigms and concerns.

Digital documents and academic productivity

What is the nature of digital documents as an innovation, that it is possible to ask whether they might affect the value of information and its use, and the organization of academic research? Geoffrey Nunberg has identified two differences between digital and mechanical technologies which affect both the value of knowledge and the organization of its reproduction.[7]

> ... unlike mechanical antecedents like the printing press, the typewriter, or the telegraph, the computer isn't restricted to a single role in production and diffusion. In fact, the technology tends to erase distinctions between the separate processes of creation, reproduction and distribution that characterize the classical industrial model of print commodities, not just because the electronic technology employed is the same at each stage, but because control over the processes can be exercised at any point. ...The second

important difference between the two technologies follows from the immateriality of electronic representations and the resulting reductions in the cost of reproduction.

The fundamental consequence of these differences, Nunberg argues, is that the user[**] has much greater control of the process of digital reproduction of knowledge as well as its content, essentially transforming the meaning of the word publication by allowing for individual customization of knowledge.

However, these differences in the process of the reproduction of ideas do not apply to every electronic document, only to true "digital documents."[8] Today's marketplace consists largely of *digitized* documents, that is, works written for and reproduced in printed journals, then scanned and distributed on the network. Digitized documents conform to the modes of production of print journals: to the rhetorical rules of the genre of scientific and to the traditional relationships between author, publisher and reader. If prior examples of technological innovation hold in this case, however, digitized documents represent only a transitional stage, one in which the attempt is made to focus the use of new technologies upon increasing the productivity of traditional modes of production and to reinforce traditional authority patterns and economic interests. CD-ROM technology is a good example of the attempt to preserve the traditional modes of production, yet take advantage of the capability of digital signals to include multimedia, by packaging them within a physical medium which behaves just like a printed commodity. The immateriality of networked information is much more difficult to control, although encryption and digital watermarking are an attempt to transform digital signals into commodity by giving an electronic signal some of the characteristics which regulate physical commodities.

The interesting points to watch will be to see if the content of digital and print versions of the same works begin to diverge, and whether readers will begin to be allowed to appropriate published works and reuse them in new contexts. Markets are made by consumers, not by publishers, and the fundamental question concerns the future of reader's behavior as the consumers of information. For example, what will be the unit of knowledge: Will readers want to consume digital journals by subscription? Or consume single articles and pay for them as stand alone commodities through document delivery? Or treat a journal run as a database and pay for access to it as a searchable information service? As Nunberg points out, the intersection of technology and markets will be determined by the nature of the digital signal, which unifies the processes of production, reproduction and use of information.

In thinking about the nature of digital documents and the kind of social relationships which they make possible, consider the impact of what may well be the most successful digital document thus far, the credit card. The credit card itself is only an interface to liquid cash and credit, taking advantage of mainframe computer technology and computer networks to manage market transactions wherever they occur around the world. It replaces printed currency, and portable forms of wealth such as letters of credit and traveler's checks, with a utility service. It creates new kinds of value: liquidity, through an interface to a world wide financial system; timeliness and access, through twenty-four hour service anywhere in the world; and customized or personalized service, through credit. These new kinds of value are not easily measured by traditional measures of productivity; Brynjolfsson notes that by traditional measures the ATM seems to reduce productivity, by reducing the use of checks, the traditional output measure of banks. Yet it is not a sufficient description of the value of credit or debit cards to characterize the new kinds of value simply as improvements in the quality of service, since they have created entirely new kinds of markets for financial services and a new interface for economic activity which supports new more mobile life styles.

One of these new markets is worthy of a second look, not only as an example of innovation, but to explore the *reflexive* quality of digital documents. When I use a debit card, a profile of my patterns of consumption is created, information which is of economic value for advertising and marketing; thus coupons for new or competing products appear on the back of my receipt. Information about my use of information is a new kind of economic value, and the basis of a new kind of market when used by advertisers and market analysts. In tracking the use of digital services, network technologies might also be described as keeping the consumer under surveillance. Issues of privacy aside, and they are not sufficiency recognized as yet, this will make possible an entirely new, direct, and unmediated relationship between consumer and publisher.

Thus the discussion of protecting intellectual property on the Internet has focused not only on technologies which allow for the control of access to copyrighted material, but also on technologies which audit the use of information, including requirements for the authentication of the identity of the user and tracking patterns of use. The consequences of this reflexivity may well reflect a fundamental shift in the way in which we conceive of the value of information. While markets for physical commodities were regulated by laws and inventory management techniques, markets for digital services will focus both upon the content and use of information, and will use the network as a medium for knowledge management techniques.[9]

To summarize this process of innovation: credit cards might be described in productivity terms as an efficient new way to manage money, but they might also be described as creating entirely new genres of wealth, literally a new kind of currency; as new ways of life which create new kinds of social and geographical mobility; and in terms of the new kinds of markets and organizations which they make possible. Digitized documents may lower the costs of reproduction and distribution of print journals, and perhaps some first copy costs, but they also create new kinds of value in faster modes of access to information, new techniques for searching, and more customized content. And in the longer run, true digital documents will produce new genres of scholarly discourse, new kinds of information markets, and perhaps new kinds of educational institutions to use them.

At the moment these new possibilities tend to be discussed in terms of the capacity of the new technology to disrupt the laws, cultures and organizations which have managed research, reading, publishing and intellectual property in the era of print. Most prominent among these has been the discussion of the protection of copy-right on the Internet, but there is also active concern about the social impacts of digital documents. There is the problem of privacy and surveillance on the Internet, particularly in the workplace. Pornography on the Web has been defined as a social problem involving the protection of children, but this is only one example of a broader issue concerning the impact of a global communications medium whose scope transcends even national regulatory authorities upon local norms and culture. And there is interest in the quality of social relationships in Cyberia, manifested negatively in the problem of hostile electronic mail, and manifested positively by emerging forms of virtual community.[10] And there is a debate in national information policy about the proper balance between the public interest in access to information, and the commercialization of information in order to create robust information markets.

Digital documents and academic community

In an essay called "The Social Life of Documents," John Seeley Brown and Paul Duguid have suggested that documents should not be understood solely as containers for content, but as catalysts for the creation of a sense of community. They say: "the circulation of documents first helps make and then helps maintain social communities and institutions in ways that looking at the content alone cannot explain. In offering an alternative to the notion that documents deliver meaning, [there is a] connection between the creation of communities and the creation of meaning."[11] That is, our attention should not be on the artifact itself, nor perhaps is the market the fundamental social formation around documents, but documents and markets create and sustain the social worlds, or communities, of readers. Here we return, at last, to the missing subject of this discussion of the causality of technology and digital documents, the academic community.

More recently, the business management literature has begun to consider an interesting variant of this thesis, that the formation of virtual communities is the most important medium for the creation and sustenance of markets for digital services. For example, John Hagel III and Arthur G. Armstrong argue that producers of digital services must adapt to the communitarian culture of the network, for,

> ...by giving customers the ability to interact with each other as well as with the company itself, businesses can build new and deeper relationships with customers. We believe that commercial success in the on-line arena will belong to those who organize virtual communities to meet multiple social and commercial needs.[12]

While producers controlled traditional markets, they argue, the information revolution shifts the balance of power to the consumer by providing tools to select the best value, creating entirely new modes of competition. The markets of the future will take the form of virtual communities which will be a medium for "direct channels of communication between producers and customers," and which will "threaten the long term viability of traditional intermediaries."[13]

The questions concerning technological innovation might now be reconstituted as a kind of sociology of knowledge: What kind of academic community first created print genres, and was in turn sustained by them? What kind of community is now creating digital genres, and is in turn sustained by them? And what is the relationship between the two, now and in the future?

On a larger scale, the relationship between virtual community and digital documents is a tacit dimension of national information policies. These are the kinds of questions that worries the People's Republic of China about creating a digital library, for the Internet is a medium for political dissent and organization, and the digital library provides access to information which has the potential to transform the scope and nature of political discourse and thereby the form of political authority. In the United States, national information policy has tended to focus on the creation of information markets, but the broader discussion of the social and political impact of digital communications has been concerned with issues of community. For example, the Communications Decency Act and subsequent judicial review has concentrated upon Internet pornography and its impact upon the culture and mores of local communities. Social and political movements ranging from Greenpeace to Militia movements have used the Internet to organize dissent and political action; is this protected free speech? Universities are concerned about the impact of abusive electronic mail upon academic culture. In each case, digital information is changing the nature of culture.

The bridge between technology and community is suggested by the elements in the analysis of productivity: how new technologies add new value, create new incentives, and enable new kinds of organization. Brown and Duguid argue that our nation's sense of political community was created by newspapers, not so much in the content of the stories, but in,

> reaching a significant portion of the population, newspapers helped develop an implicit sense of community among the diverse and scattered populace of the separate colonies and the emerging post-revolutionary nation...That is, the emergence of a common sense of community contributed as much to the formation of nationhood as the rational arguments of Common Sense. Indeed the former helped create the audience for the latter.[14]

Similarly, and closer to the issue of scholarly communication, the scientific letters which circulated among the Fellows of the Royal Society were the prototype for scientific journals, which in turn sustained scholarly disciplines, which are the organizing infrastructure for academic literature and departments. Let us postulate then, for heuristic purposes, that when we speak of added value, we are beginning a discussion of the process of innovation which is linked to an understanding of community formation. New forms of value, which is to say new forms for the *use* of information, create new genres of documents, which in turn create a literature, which serves as the catalyst and historical memory for new forms of communities.

In the case of print and digital documents, change is not evolutionary because these two kinds of information offer different kinds of value, but neither are they opposites. Genre, for example, has been shaped by the physical characteristics of the print medium, including the *design* of information (e.g., page layout, font), as well as the rhetorical norms governing the structure of information (e.g., essay, scientific article, novel). Rhetoric has been described as a strategy for managing the allocation of the scarcest resource of modern times, our attention. Although we often complain of "information overload," this may well reflect an early stage in the development of rhetorical structures for modern media. Certainly there is more information, and more kinds of information, but the real problem is the difficulty in determining the quality of digital information (e.g., the lack of reputation and branding); the difficulty of knowing which kind of information is relevant for certain kinds of decisions (e.g., the problem of productivity); and the relatively primitive rhetorical rules which govern new media (e.g., the problem of flaming in electronic mail).

Consider, for example, the technology of scientific visualization and multimedia. Thus far, visual culture has been governed largely by the rhetorical rules of entertainment, which require us to surrender our critical judgment in order to enjoy the show. Thus the problem of the quality of multimedia information is not simply technical, but requires the development of new genres and rhetorical norms within which visual media are consistent with academic values such as critical judgment.

Or, consider some of the new genres for digital documents, which might well be described as adding new kinds of value to information: hypertext, the Boolean search, and the database. The database raises new questions about the unit of knowledge, as we have seen. Will consumers subscribe to and read the digital *journal*, pay for network delivery of digital *articles*, or will the unit of knowledge be the *screen*, the digital analog of the paragraph, which is identified by a search engine or agent? HTML raises the question: who is responsible for the context of information, the author or the reader? If one can jump from text to text, linking things which had not previously been linked, it is the reader who creates context and therefore governs meaning, and reading becomes a kind of performing art.

These questions might be described, perhaps, as a legitimation crisis, in that the traditional authorities which governed or mediated the structure and quality of print are no longer authoritative: the author, the editor, the publisher and the library. Who are the new authorities?

Sociologically, there is no doubt that the information problems of engineers and scientists were the template from which new genres and rhetorical forms evolved, becoming instantiated into hardware and software, thence into computer literacy, and thence the user skills and modes of reading or using information. Hypertext, for example, turns narrative into a database, which is a highly functional strategy for recovering specific bits of information in scientific research, as, for example, in searching for information with which to solve a problem. Electronic mail is a highly efficient means for exchanging messages but has little scope for lexical or rhetorical nuance. This is not a problem for groups sharing a common culture and background, like scientists and engineers, but it becomes a problem given the diverse social groups which use electronic mail as a medium for communication today, hence the frequency of flaming and misunderstanding.

As sociologists like Bruno Latour have noted, in any case, the original intent of the designers of a technology does not necessarily govern the process of technological innovation, for the meaning and purpose of a technology mutates as it crosses social contexts.[15] Thus the problem is not best posed in terms of the cultural hegemony of the sciences and technology over academic institutions, but these origins can still be recognized when we give "commands" to a computer.

But there is an interesting problem to be thought about, namely the cultural and organizational consequences of information technologies which originated in other sectors of the economy, from business and the military, for the academic community. Thus far the discussion of this topic has occurred at the boundary of the academic enterprise, often in the context of thinking about the uses of distance education, which is to say, the extension of the scope of a given institution's teaching services to a national, or perhaps global, market. But there is a broader question about the nature of the academic community itself in a research University: what is the substance of this sense of community, and what sustains it?

While it is often claimed that digital communication can sustain a sense of *virtual* community, what is meant by virtual, and what is meant by community? The literature on *social capital* argues that civic virtue is a function of participation, and those who participate in one voluntary social activity are highly likely to participate in others, creating a social resource called civil society or community.[16] Robert Putnam argues that television, and perhaps other media, are a passive sort of participation which replace and diminish civic communities. The question is whether today's virtual communities represent a kind of social withdrawal, or whether they might come to be resources for social participation and community. If this is an important goal of digital networks, how can they be designed to this purpose. Can networks be designed to facilitate the moral virtues of community, such as trust, reciprocity, and loyalty?

And finally, to return to the question of the productivity of knowledge workers in an information society, and mindful of the heuristic principle that documents can be understood in terms of the communities they sustain, is not the research library best conceptualized as the traditional knowledge management strategy of the academic community? If so, how well does the digital library perform this function, at least as we understand it thus far? Other than the parking lot, perhaps, the Library is one of the last public or common goods in an academic world which is increasingly specialized, and perhaps fragmented. The digital library, however, is generally conceived of only as an information resource, as if the Library were only the container for a collection, rather than a shared intellectual resource and site for a community.

The social functions of the Library are not easily measured in terms of outcomes, but are an element in the productivity of faculty and students. To some extent, perhaps, libraries have brought this problem on themselves by measuring their quality in terms of fiscal inputs and size of collections, and must begin to define and measure their role in productivity and community formation. But in another sense, the focus upon the content and format of information to the exclusion of consideration of the social contexts and functions of knowledge is a distortion of the nature and dynamics of scholarly communication and the academic community.

NOTES:

[*] This paper is the written version of an after dinner speech at the conference on Scholarly Communication and Technology, presented on April 24, 1997.

[1] Peter F. Drucker, **The Age of Discontinuity** (New York: Harper and Row, 1978)

[2] Walter Wriston, **The Twilight of Sovereignty** (New York: Charles Scribner's Sons, 1992), xii.

[3] Erik Brynjolfsson, "The Productivity Paradox of Information Technology: Review and Assessment." **Communications of the ACM** (December, 1993), 36(12)67-77.

[4] See, for example, William F. Massy and Robert Zemsky, "Using Information Technology to Enhance Academic Productivity," a White Paper for the EDUCOM National Learning Infrastructure Initiative (NLII), 1995.

[5] Anthony M. Cummings, Marcia L. Witte, William G. Bowen, Laura O. Lazarus, and Richard H. Ekman. **University Libraries and Scholarly Communication**. Washington DC: The Association of Research Libraries, 1992.

[6] See Manuel Castells, **The Rise of the Network Society** (London: Blackwell Publishers, 1996) 56.

[7] Geoffrey Nunberg, "The Places of Books in the Age of Electronic Reproduction," **Future Libraries**. R. Howard Bloch and Carla Hesse, editors. (Berkeley: University of California Press, 1993) 21-22.

[**] The term "user" is applied to the consumption of digital documents, the way the term "reader" is applied to the consumption of printed works. "User" is a kind of semantic strategy

for pointing out that engagement with a technology mediates between the reader and the text, allowing for the direct control over content and format which Nunberg describes.

[8] See, for example, the report on a workshop held at the Hawaii International Conference on System Sciences conference in January 1995, edited by M. Stuart Lynn and Ralph H. Sprague, Jr.: Documents in the Digital Culture: Shaping the Future. Honolulu, Hawaii: A HICSS Monograph, 1995.

[9] On knowledge management, see, for example: Thomas H. Davenport, Sirkka L. Jarvenpaa and Michael C. Beers, "Improving Knowledge Work Processes," **Sloan Management Review** (Summer 1996), 53-65.

[10] The term Cyberia reflects the anthropologist's approach to analyzing the Internet as a site for culture and community, best summarized by Arturo Escobar, "Welcome to Cyberia: Notes on the Anthropology of Cyberculture." **Current Anthropology** (June 1994), 35(4)211-231.

[11] John Seeley Brown and Paul Duguid, "The Social Life of Documents," **Release 1.0: Esther Dyson's Monthly Report.** (New York: Edventure Holdings Inc., October 11, 1995) 7.

[12] John Hagel III and Arthur G. Armstrong, **net.gain: expanding markets through virtual communities**. (Boston: Massachusetts: Harvard Business School Press, 1997)5.

[13] **Ibid**., 204

[14] Brown and Duguid, **op.cit.**, 5. This argument is derived from Benedict Anderson's **Imagined Communities.**

[15] Bruno Latour, **Aramis, or The Love of Technology.** Translated by Catherine Porter. (Cambridge, Mass: Harvard University Press, 1996). Note that Latour chose the novel genre to discuss this phenomenon.

[16] See, for example: Robert D. Putnam, "The Strange Disappearance of Civic America," **The American Prospect** (Winter 1996), 24(34-48). In the same issue, see also: Sherry Turkle, "Virtuality and Its Discontents: Searching for Community in Cyberspace."

◆ ◆ ◆

For additional information about the conference, or The Andrew W. Mellon Foundation's scholarly communication initiatives, please contact Richard Ekman. For additional information about ARL or this web site contact Patricia Brennan, ARL Program Officer at (202) 296-2296.

Return to Office of Scholarly Communication Home Page

ARL Home ARL Scholarly Communication and Technology Home Page

© Association of Research Libraries, Washington, DC

Web Design by Angelo F. Cruz
Maintained by ARL Web Administrator
Last Modified: August 26, 1997

Scholarly Communication and Technology

Conference Organized by The Andrew W. Mellon Foundation

at Emory University
April 24-25, 1997

Copyright © of the papers on this site are held by the individual authors or The Andrew W. Mellon Foundation. Permission is granted to reproduce and distribute copies of these works for nonprofit educational or library purposes, provided that the author, source, and copyright notice are included on each copy. For commercial use, please contact **Richard Ekman** at the The Andrew W. Mellon Foundation.

Session #5 Technical Choices and Standards

Making Technology Work for Scholarship: Investing in the Data

Susan Hockey
Department of English
University of Alberta

. . . .

Making Technology Work for Scholarship: Investing in the Data

Susan Hockey
Department of English
3-5 Humanities Centre
University of Alberta
Edmonton
Alberta
T6G 2E5
Canada
Phone: 403.492.1029
Fax: 403.492.8142
E-mail: Susan.Hockey@UAlberta.ca

Revised May 1997

The introduction of any kind of new technology is often a painful and time-consuming process, at least for those who must incorporate it into their everyday lives. This is particularly true of computing technology where the learning curve can be steep, what is learned changes rapidly and ever more new and exciting things seem to be perpetually on the horizon. How can the providers and consumers of electronic information make the best use of this new medium and ensure that the information they create and use will outlast the current system on which it is used? In this paper we examine some of these issues, concentrating on the humanities where the nature of the information studied by scholars can be almost anything and where the information can be studied for almost any purpose.

Today's computer programs are not sophisticated enough to process raw data sensibly. This situation will remain true until artificial intelligence and natural language processing research has made very much more progress than it has so far. Very early on in my days as a humanities computing specialist I saw a library catalogue which had been typed into the computer without anything to separate the fields in the information. There was no way of knowing what was the author, title, publisher or call number of any of the items. The catalogue could be printed out but the titles could not be searched at all, nor could the items in the catalogue be sorted by author name. Although a human can tell which is the author or title from reading the catalogue, a computer program cannot. Something must be inserted in the data to give the program more information. This is a very simple example of markup or encoding which is needed to make computers work better for us. Since we are so far off having the kind of intelligence we really need in computer programs, we must put that intelligence in the data so that computer programs can be informed by it. The more intelligence there is in our data, the better our programs will perform. But what should that intelligence look like? How can we ensure that we make the right decisions in creating it so that computers can really do what we want? Some scholarly communication and digital library projects are among those which are beginning to provide answers to these questions.

1. New Technology or Old?

That having been said, we see many current technology and digital library projects concentrating on using the new technology as an access mechanism to deliver th e old technology. They assume that the typical scholarly product is an article or monograph and that it will be read in a sequential fashion as indeed we have done for hundreds of years ever since these products began to be produced on paper and bound into physical artefacts such as books. The difference is only that instead of going to the library or bookstore to obtain the object, we access it over the network - and then almost certainly have to print a copy of it in order to read it. Of course there are tremendous savings of time for those who have instant access to the network, can find the material they are looking for easily and have high-speed printers. I want to argue here that delivering the old technology via the new is only a transitory phase and that it must not be viewed as an end in itself. Before we embark on the large-scale compilation of electronic information, we must consider how future scholars might use this information and what are the best ways of ensuring that the information will last beyond the current technology.

The old (print) technology developed into a sophisticated model over a long period of time.[1]

Books consist of pages which are bound up in sequential fashion, delivering the text in a single linear sequence. Page numbers and running heads are used for identification purposes. Books also often include other organizational aids such as tables of contents and back-of-the book indexes which are conventionally placed at the beginning and end of the book. Footnotes, bibliographies, illustrations etc provide additional methods of cross-referencing. A title page provides a convention for identifying the book and its author and publication details. The length of a book is often determined by publishers' costs or requirements, rather than by what the author really wants to say about the subject. Journal articles also exhibit similar characteristics, being also designed for reproduction on pieces of paper. Furthermore, the ease of reading of printed books and journals is determined by their typography which is designed to help the reader by reinforcing what the author wants to say. Conventions of typography (headings, italic, bold etc) make things stand out on the page for the human eye.

When we put information into electronic form, we find that we can do many more things with it than we can with a printed book. We can still read it, though not as well as we can read a printed book. The real advantage of the electronic medium is that we can search and manipulate the information in many different ways. We are no longer dependent on the back-of-the-book index to find things in the information but can search for any word or phrase using retrieval software. We no longer need the whole book to look up one paragraph in it, but can just access the piece of information we need. We can also access several different pieces of information at the same time and make links between them. We can find a bibliographic reference and go immediately to the place to which it points. We can merge different representations of the same material into a coherent whole and we can count instances of features within the information. We can thus begin to think of the material we want as "information objects".[2]

To reinforce the arguments we are making here, we can call electronic images of printed pages "dead text" and use the term "live text" for searchable representations of text.[3] For dead text we can use only those retrieval tools which were designed for finding printed items and even then this information must be added as searchable live text, usually in the form of bibliographic references or tables of contents. Of course most of the dead text produced over the last fifteen or so years began its life as live text in the form of wordprocessor documents. The obvious question is how can the utility of that live text be retained and not be lost for ever.

2. Electronic Text and Data Formats

Long before digital libraries became popular, live electronic text was being created for many different purposes, most often, as we have seen, with word processing or typesetting programs. Unfortunately this kind of live electronic text is normally only searchable by the word processing program which produced it and then only in a very simple way. We have all encountered the problems involved in moving from one word processing program to another. Although some of these problems have been solved in more recent versions of the software, maintaining an electronic document as a word processing file is not a sensible option for the long term, unless the creator of the document is absolutely sure that this document will only ever be needed in the short-term future and only ever for the purposes of word processing by the program that created it. Word processed documents contain typographic markup or codes to specify the formatting. If there was no markup the document would be much more difficult to read. However typesetting markup is ambiguous and thus cannot be used sensibly by any retrieval program. For example, italics can be used for titles of books, or for emphasized words, or for foreign words. With typographic markup we cannot distinguish titles of books from foreign words, which we may at some stage want to search for separately.

Other electronic texts were created for the purposes of retrieval and analysis. Many such examples exist, ranging from the large text databases of legal statutes to humanities collections such as the Thesaurus Linguae Graecae (TLG) and the Trésor de la langue française. These projects all realized that they needed to put some intelligence into the data in order to search it effectively. Most devised markup schemes which focus on ways of identifying the reference citations for items which have been retrieved, for example in the TLG, the name of the author, work, book and chapter number. They do not provide easily for representing items of interest within a text, for example foreign words or quotations. Most of these markup schemes are specific to one or two computer programs and texts prepared in them are not easily interchangeable. A meeting in 1987 examined the very many markup schemes for humanities electronic texts and concluded that the present situation was "chaos".[4] No existing markup scheme satisfied the needs of all users and much time was being wasted converting from one deficient scheme to another.

Another commonly used method of storing and retrieving information is a relational database, as, for example, in Microsoft Access or dBASE, or the mainframe program Oracle. In it, data is assumed to take the form of one or more tables consisting of rows and columns, that is rectangular structures.[5] A simple table of biographical information may have rows representing people and columns holding information about those people, for example, name, date of birth, occupation etc. When a person has more than one occupation, the data becomes clumsy and the information is best represented in two tables where the second has a row for each person's occupation. The tables are linked or related by the person. A third table may hold information about the occupations. It is not difficult for a human to conceptualize the data structures of a relational database or for a computer to process them. Relational databases work well for some kinds of information, for example address lists etc, but in reality not much data in the real world fits well into rectangular structures. This means that the information is distorted when it is entered into the computer, and processing and analyses are carried out on the distorted forms, whose distortion tends to be forgotten. Relational databases also force the allocation of information to fixed data categories, whereas, in the humanities at any rate, much of the information is subject to scholarly debate and dispute, requiring multiple views of the material to be represented. Furthermore, getting information out of a relational database for use by other programs usually requires some programming knowledge.

The progress of too many retrieval and database projects can be characterized as follows. The project decided that it wants to "make a CD-ROM". It finds that it has to investigate possible software programs for delivery of the results and chooses the one which has the most seductive user interface or most persuasive salesperson. If the data includes some non-standard characters, being able to display them on the screen is considered the highest priority and the functions that are needed to manipulate those characters are not looked at very hard. Data is then entered directly into this software over a period of time during which the software interface begins to look outmoded as technology changes. By the time that the project has finished entering the data, the software company has gone out of business leaving the project with a lot of valuable information in a proprietary software format which is no longer supported. More often than not the data is lost and much time and money has been wasted. The investment is clearly in the data and it makes sense to ensure that this is not dependent on one particular program, but can be used by other programs as well.

3. The Standard Generalized Markup Language (SGML)

Given the time and effort involved in creating electronic information, it makes sense to step back and think about how to ensure that the information can outlast the computer system on which it is created, and can also be used for many different purposes. These are the two main principles of the Standard Generalized Markup Language (SGML) which became an international standard (ISO 8879) in 1986.[6] SGML was designed as a general purpose markup scheme that can be applied to many different types of documents and in fact to any electronic information. It consists of plain ASCII files which can easily be moved from one computer system to another. SGML is a descriptive language. Most encoding schemes prior to SGML use prescriptive markup. One example of prescriptive markup is word processing or typesetting codes embedded in a text which give instructions to the computer such as "center the next line" or "print these words in italic". Another is fielded data which is specific to a retrieval program, for example, reference citations or author's names which must be in a specific format for the retrieval program to recognize them as such. By contrast, a descriptive markup language merely identifies what the components of a document are. It does not give specific instructions to any program. In it, for example, a title is encoded as a title, or a paragraph as a paragraph. This very simple approach ultimately allows much more flexibility. A printing program can print all the titles in italic. A retrieval program can search on the titles and a hypertext program can link to and from the titles, all without making any changes to the data.

Strictly, SGML itself is not a markup scheme, but a kind of computer language for defining markup or encoding schemes. SGML markup schemes assume that each document consists of a collection of objects which nest within each other or are related to each other in some other way. These objects or features can be almost anything. Typically they are structural components such as title, chapter, paragraph, heading, act, scene, speech, but they can also be interpretive information such as parts of speech, names of people and places, quotations (direct and indirect) and even literary or historical interpretation. The first stage of any SGML-based project is document analysis where the project identifies all the textual features which are of interest and the relationships between them. This can take some time, but it is worth investing the time since a thorough document analysis can ensure that data entry proceeds smoothly and that the documents are easily processable by computer programs.

In SGML terms, the objects within a document are called elements. They are identified by a start and end tag as follows: <title>Pride and Prejudice</title>. The SGML syntax allows the document designer to specify all the possible elements as a Document Type Declaration (DTD) which is a kind of formal model of the document structure. The DTD indicates which elements are contained within other elements, which are optional, which can be repeated etc. For example, in simple terms a journal article consists of a title, one or more authors, an optional abstract, an optional list of keywords, followed by the body of the article. The body may contain sections, each beginning with a heading followed by one or more paragraphs of text. The article may finish with a bibliography. The paragraphs of text may contain other features of interest including quotations, lists, names, as well as links to notes. A play has a rather different structure of which an outline could be: title, author, castlist, one or more acts each containing one or more scenes, each containing one or more speeches and stage directions etc.

SGML elements may also have attributes which further specify or modify the element. One use of attributes may be to normalize the spelling of names for indexing purposes. For example, the name Jack Smyth could be encoded as <name norm="SmithJ"> Jack Smyth</name>, but indexed under S as if it were Smith. Attributes can also be used to normalize date forms for sorting, for example <date norm=19970315>the Ides of March 1997</date>. Another important function of attributes is to assign a unique identifier to each instance of each SGML element within a document. This can be used as a cross-reference by any kind of hypertext program. The list of possible attributes for an element may be defined as a closed set, allowing the encoder to pick from a list, or it may be entirely open.

SGML has another very useful feature. Any piece of information can be given a name and be referred to by that name in an SGML document. These are called entities and are enclosed in & and ;. One use is for non-standard characters, where for example é can be encoded as &eacute; thus ensuring that it can be transmitted easily across networks and from one machine to another. A standard list of these characters exists, but the document encoder can also create more. Entity references can also be used for any boilerplate text. This avoids repetitive typing of words and phrases which are repeated, thus also reducing the chance of errors. An entity reference can be resolved to any amount of text from a single letter up to something like a whole chapter.

The formal structure of SGML means that the encoding of a document can be validated automatically, a process known as parsing. The parser makes use of the SGML DTD to determine the structure of the document and can thus help to eliminate whole classes of encoding errors, before the document is processed by an application program. For example, an error can be detected if the DTD specifies that a journal article must have one or more authors, but the author's name has been omitted accidentally. Mistyped element names can be detected as errors as can elements which are wrongly nested, for example, an act within a scene when the DTD specifies that acts contain scenes. Attributes can also be validated when there is a closed set of possible values. The validation process can also detect un-resolved cross-references which use SGML's inbuilt identifiers. The SGML document structure and validation process means that any application program can operate more efficiently since it derives information from the DTD about what to expect in the document. It follows that the stricter the DTD, the easier it is to process the document. However very strict DTDs may force the document encoder to make decisions which simplify what is being encoded. Free DTDs might better reflect the nature of the information but usually require more processing. Another advantage of SGML is very apparent here. Once a project is underway, if a document encoder finds a new feature of interest, that feature can simply be added to the DTD without the need to restructure work that has already been done. Of course many documents can be encoded and processed with the same DTD.

4. The Text Encoding Initiative

The humanities computing community was among the early adopters of SGML, for two very simple reasons. Humanities primary source texts can be very complex, and they need to be shared and used by different scholars. They can be in different languages and writing systems and can contain textual variants, non-standard characters, annotations and emendations, multiple parallel texts, hypertext links, as well as having complex canonical reference systems. In electronic form, these texts can be used for many different purposes including the preparation of new editions, word and phrase searches, stylistic analyses and research on syntax and other linguistic features. By 1987 it was clear that many encoding schemes existed for humanities electronic texts, but none was sufficiently powerful to allow for all the different features which might be of interest. Following a planning meeting at which representatives of leading humanities computing projects were present, a major international project called the Text Encoding Initiative (TEI),was launched.[7] Sponsored by the Association for Computers and the

Humanities, the Association for Computational Linguistics and the Association for Literary and Linguistic Computing, the TEI enlisted the help of volunteers all over the world to define what features might be of interest to humanities scholars working with electronic text. It built on the expertise of groups such as the Perseus Project (then at Harvard, now at Tufts University), the Brown University Women Writers Project, the Alfa Informatica Group in Groningen, Netherlands, and others who were already working with SGML, to create SGML tags which could be used for many different types of text.

The TEI published its *Guidelines for the Encoding and Interchange of Electronic Texts*, in May 1994 after over six years' work. The Guidelines identify some four hundred tags, but of course no list of tags can be truly comprehensive and so the TEI builds up its DTDs in a way which makes it easy for users to modify them. The TEI SGML application is built on the assumption that all text share some common core of features to which can be added tags for specific application areas. Very few tags are mandatory and most of these are concerned with documenting the text and will be discussed further below. The TEI Guidelines are simply guidelines. They serve to help the encoder identify features of interest and provide the DTDs with which the encoder will work. The core consists of the header which documents the text, plus basic structural tags and common

features such as lists, abbreviations, bibliographic citations, quotations, simple names and dates etc. The user selects a base tag set of which the following have been defined at present: prose, verse, drama, dictionaries, spoken texts, terminological data. To this are added one or more additional tag sets. The options here include simple analytic mechanisms, linking and hypertext, transcription of primary sources, critical apparatus, names and dates, and some methods of handling graphics. The TEI has also defined a method of handling non-standard alphabets by using a Writing System Declaration which the user specifies. It can also be used for non-alphabetic writing systems, for example, Japanese, Building a TEI DTD has thus been likened to the preparation of a pizza where the base tag set is the base, the core tags are the tomato and cheese and the additional tag sets are the toppings.

One of the issues addressed at the TEI planning meeting was the need for documentation of an electronic text. Many electronic texts now exist about which little is known, either what source text they were taken from, what decisions were made in encoding the text and what changes have been made to the text. All this information is extremely important to a scholar wanting to work on the text, since it will determine the academic credibility of his or her work. Unknown sources are unreliable at best and lead to inferior work. Experience has shown that electronic texts are more likely to contain errors or have bits missing, but these are more difficult to detect than with printed material. It seems that one of the main reasons for this lack of documentation for electronic texts was simply that there was no common methodology for providing it.

The TEI examined various models for documenting electronic texts and concluded that some SGML elements placed as a header at the beginning of an electronic text file would be the most appropriate way of providing this information. Since the header is part of the electronic text file, it is more likely to remain with that file throughout its life. It can also be processed by the same software as the rest of the text. The TEI header contains four major sections.[8] One is a bibliographic description of the electronic text file using SGML elements which map closely on to some MARC fields. The electronic text is a different intellectual object from the source from which it was created and the source is thus also identified in the header. The encoding description section provides information about the principles used in encoding the text, for example whether the spelling has been normalized, treatment of end-of-line hyphens, etc. For spoken texts the header provides a way of identifying the participants in a conversation and attaching a simple identifier to each of them which can then be used as an attribute on each utterance. The header also provides a revision history of the text indicating who made what changes to it and when.

As far as can be ascertained the TEI header is the first systematic attempt to provide documentation for an electronic text a part of the text file itself. A good many projects are now using it, but experience has shown that it would perhaps benefit from some revision. Scholars find it hard to create good headers. Some elements in the header are very obvious, but the relative importance of the remaining elements is not so clear. At some institutions librarians are creating TEI headers, but they need training in the use and importance of the non-bibliographic sections and in how the header can be used by computer software other the bibliographic tools which they know well.

5. Encoded Archival Description (EAD)

Another SGML application which has attracted a lot of attention in the scholarly community and archival world is the Encoded Archival Description (EAD). First developed by Daniel Pitti at the University of California at Berkeley and now taken over by the Library of Congress, the EAD is an SGML application for archival finding aids.[9] Finding aids are very suitable for SGML because they are basically hierarchic in structure. In simple terms a collection is divided into series which consist of boxes which contain folders etc. Prior to the EAD, there was no effective standard way of preparing finding aids. Typical projects created a collection level record in one of the bibliographic utilities such as RLIN and used their own procedures, often a word processing program, for creating the finding aid. Possibilities now exist for using SGML to link electronic finding aids with electronic representations of the archival material itself. One such experiment, conducted at the Center for Electronic Texts in the Humanities (CETH), has created an EAD-encoded finding aid for part of the Griffis Collection at Rutgers University and encoded a small number of the items in the collection (19th century essays) in the TEI scheme.[10] The user can work with the finding aid to locate the item of interest and then move directly to the encoded text and an image of the text to study the item in more detail. The SGML browser program Panorama allows the two DTDs to exist side by side and in fact uses an extended pointer mechanism devised by the TEI to move from one to the other.

6. Other Applications of SGML

SGML is now being widely adopted in the commercial world as companies see the advantage of investment in data which will move easily from one computer system to another. It is worth noting that the few books on SGML which appeared early in its life where intended for an academic audience. More recent books are intended for a commercial audience and emphasize the cost savings involved in SGML as well as the technical requirements. This is not to say that these books are not of any value to academic users. The SGML Web pages list many projects in the areas of health, legal documents, electronic journals, rail and air transport, semiconductors, the US Internal Revenue Service and more. SGML is extremely useful for technical documentation as can be evidenced by the list of customers on the Web page of one of the major SGML software companies INSO/EBT. This includes United Airlines, Novell, British Telecom, AT&T, Shell, Boeing, Nissan and Volvo.

SGML need not only be used with textual data. It can be used to describe almost anything. SGML should not therefore be seen as an alternative to Acrobat, PostScript or other document formats, but as a way of describing and linking together documents in these and other formats, forming the "underground tunnels" which make the documents work for users.[11] SGML can be used to encode the searchable textual information which must accompany images or other formats in order to make them useful. With SGML the searchable elements can be defined to fit the data exactly and can be used by different systems. This is in contrast with storing image data in some proprietary database system, as often happens. Further down the line we can imagine a situation where a scholar wants to examine the digital image of a manuscript and also have available a searchable text. He or she may well find something of interest on the image and want to go to occurrences of the same feature elsewhere within the text. In order to do this, the encoded version of the text must know what that feature of interest is and where it occurs on the digital image. Knowing which page it is on is not enough. The exact position on the page must be encoded. This information can be represented in SGML which thus provides the sophisticated kind of linking needed for scholarly applications. SGML structures can also point to places within a recording of speech or other sound and can be used to link the sound to a transcription of the conversation, again enabling the sound and text to be studied together. Other programs exist which can perform these functions, but the problem with all of them is that they use a proprietary data format which cannot be used for any other purpose.

7. SGML, HTML and XML

The relationship between SGML and the HyperText Markup Language (HTML) needs to be clearly understood. Although not originally designed as such, HTML is now an SGML application, even though many HTML documents exist which cannot be validated according to the rules of SGML. HTML consists of a set of elements which are interpreted by Web browsers for display purposes. The HTML tags were designed for display and not for other kinds of analysis, which is why only crude searches are possible on Web documents. HTML is a rather curious mixture of elements. Larger ones such as <body>, <h1> etc, <p> for paragraph, <ul> for unordered list are structural, but the smaller elements such as <b> for bold, <i> for italic are typographic, which, as we have seen above, is ambiguous and thus cannot be searched effectively. HTML version 3 attempts to rectify this somewhat by introducing a few semantic level elements, but these are very few in comparison with those identified in the TEI core set. HTML can be a good introduction to structured markup. Since it is so easy to create, many projects begin by using HTML and graduate to SGML once they have got used to working with structured text and begin to see the weakness of HTML for anything other than the display of text. SGML can easily be converted automatically to HTML for delivery on the Web, and Web clients have been written for the major SGML retrieval programs.

The move from HTML to SGML can be substantial and in 1996 work began on XML (Extensible Markup Language) which is a simplified version of SGML for delivery on the Web. It is "an extremely simple dialect of SGML" the goal of which "is to enable generic SGML to be served, received, and processed on the Web in the way that is now possible with HTML". XML is being developed under the auspices of the World Wide Web Consortium and the first draft of the specification for it was available by the SGML conference in December 1996. Essentially it is SGML with some of the more complex and esoteric features removed. It has been designed for interoperability with both SGML and HTML, to fill the gap between the HTML which is too simple and full-blown SGML which can be complicated. As yet there is no specific XML software, but the work of this group has considerable backing and the design of XML has proceeded quickly.[12]

8. SGML and New Models of Scholarship

SGML's object-like structures make it possible for scholarly communication to be seen as "chunks" of information which can be put together in different ways. Using SGML we no longer have to squeeze the product of our research into a single linear sequence of text, whose size is often determined by the physical medium in which it will appear, but can organize it in many different ways, privileging some for one audience and others for a different audience. Some projects are already exploiting this potential and I am collaborating in two which are indicative of the way I think humanities scholarship will develop in the 21st century. Both make use of SGML to create information objects which can be delivered in many different ways.

The Model Editions Partnership (MEP) is defining a set of models for electronic documentary editions.[13] Directed by David Chesnutt of the University of South Carolina with the TEI Editor, C. Michael Sperberg-McQueen, and myself as co-coordinators, the MEP also includes seven documentary editing projects. Two of these projects are creating image editions and the other five are preparing letterpress publications. These documentary editions provide the basic source material for the study of American history, by adding the historical context which makes the material meaningful to readers. Much of this source material consists of letters which often refer to people and places by words which only the author and recipient understand. A good deal is in handwriting which only scholars specializing the field can read. Documentary editors prepare the material for publication by transcribing the documents, organizing the sources into a coherent sequence which tells the story (the history) behind them, and annotating them with information to help the reader understand them. However, the printed page is not very good vehicle for conveying the information which documentary editors need to say. It forces one organizing principle on the material (the single linear sequence of the book), when the material could well be organized in several different ways (chronologically or by recipient of letters). Notes must appear at the end of an item to which they refer or at the end of the book. When the same note, for example, a short biographical sketch of somebody mentioned in the sources, is needed in several places, it can only appear once and then be cross-referenced by page numbers, often to earlier volumes. If something has been crossed out and rewritten in a source document, this can only represented clumsily in print, even though it may reflect a change of mind which altered the course of history.

At the beginning of the MEP project, the three coordinators visited all seven partner projects, showed them some very simple demonstrations and then invited them to "dream" about what they would like to do in this new medium. The ideas collected during these visits were the incorporated into a prospectus for electronic documentary editions. The MEP sees SGML as the key to providing all the functionality outlined in the prospectus. The MEP has developed an SGML DTD for documentary editions which is based on the TEI and has begun to experiment with delivery of samples from the partner projects. The material for the image editions is wrapped up in an "SGML envelope" which provides the tools to access the images. This envelope can be generated automatically from the relational databases in which the image access information is now stored. For the letterpress editions, many more possibilities are apparent. If desired, it will be possible to merge material from different projects which are working on the same period of history. It will be possible to select subsets of the material easily, by any of the features that are tagged. This means that editions for high school students or the general public could be created almost automatically from the archive of scholarly material. With a click of a mouse the user can go from a diplomatic edition to a clear reading text and thus trace the author's thoughts as the document was being written. The documentary editions also include very detailed conceptual indexes compiled by the editors. It will be possible to use these as an entry point to the text and also to merge indexes from different projects. The MEP sees the need for making "dead text" image representations of existing published editions available quickly and believes that these can be made much more useful by wrapping them in SGML and using the conceptual indexes as an entry point to them.

The second project is even more ambitious than the MEP, since it is dealing with entirely new material and has been funded for five years. The Orlando Project at the Universities of Alberta and Guelph is a major collaborative research initiative funded by the Canadian Social Sciences and Humanities Research Council.[14] Directed by Patricia Clements, the project is creating an Integrated History of Women's Writing in the British Isles, which will appear in print and electronic formats. The project has a team of graduate research assistants carrying out basic research for the project in libraries and elsewhere. The research material they are assembling is being encoded in SGML so that it can be retrieved in many different ways. SGML DTDs have been designed to reflect the biographical details for each woman writer, also their writing history, other historical events which influenced their writing, a thesaurus of keyword terms etc. The DTDs are based on the TEI but they incorporate much descriptive and interpretive information, reflecting the nature of the research and the views of the literary scholars in the team. Tagsets have been devised for topics such as the discussion of issues of authorship and attribution, for genre issues and for issues of reception of an author's work.

The Orlando Project is thus building up an SGML-encoded database of many different kinds of information about women's writing in the British Isles. The SGML encoding, for example, greatly assists in the preparation of a chronology by allowing the project to pull out all chronology items from the different documents and sort them by their dates. It facilitates an overview of where the women writers lived, their social background, what external factors influenced their writing etc. It helps the creation and consistency of new entries since the researchers can see immediately if similar information has already been encountered. The authors of the print volumes will draw on this SGML archive as they write, but the archive can also be used to create many different hypertext products for research and teaching.

Both Orlando and the MEP are essentially working with pieces of information, which can be linked in many different ways. The linking, or rather the interpretation which gives rise to the linking is essentially what humanities scholarship is about. When the information is stored as encoded pieces of information, it can be put together in many different ways and used for many different purposes of which creating a print publication is only one. We can expect other projects to begin to work in this way as they see the advantages of encoding the features of interest in their material and manipulating them in different ways.

It is useful to look briefly at some other possibilities. Dictionary publishers were among the first to use SGML. (Although not strictly SGML, since it does not have a DTD, the Oxford English Dictionary was the first academic project to use structured markup.) When well designed, the markup enables the dictionary publishers to create spin-off products for different audiences by selecting a subset of the tagged components of an entry. A similar process can be used for other kinds of reference works. Tables of contents, bibliographies, and indexes can all be compiled automatically from SGML markup and can also be cumulative across volumes or collections of material.

The MEP is just one project that uses SGML for scholarly editions. A notable example is the CD-ROM of Chaucer's Wife of Bath's Prologue prepared by Peter Robinson and published by Cambridge University Press in 1996. This CD-ROM contains all fifty-eight pre-1500 manuscripts of the text with encoding for all the variant readings, as well as digitized images of every page of all the manuscripts. Software programs provided with the CD-ROM can manipulate the material in many different ways enabling a scholar to collate manuscripts, move immediately from one manuscript to another, compare transcriptions, spellings and readings. All the material is encoded in SGML and it includes over one million hypertext links which were generated by a computer program. This means that the investment in the project's data is carried forward from one delivery system to another, indefinitely into the future.

9. Making SGML Work Effectively

Getting started with SGML can seem to be a big hurdle to overcome, but in fact the actual mechanics of working with SGML are nowhere near as difficult as is often assumed. SGML tags are rarely typed in, but are normally inserted by software programs. WordPerfect 6.1 and 7 includes an SGML component and many projects use SoftQuad's Author/Editor for data entry. These programs can incorporate a template which is filled in with data. Like other SGML software they make use of the DTD. They know which tags are valid at any position in the document and can offer only those to the user who can pick from a menu. They can also provide a pick list of attributes and their values if these are a closed set. They ensure that what is produced is a valid SGML document. They can also toggle the display of tags on and off very easily - Author/Editor and other SoftQuad products enclose them in boxes which are very easy to see. They also incorporate style sheets which define the display format for every element.

Nevertheless, inserting tags in this way can be rather cumbersome and various software tools exist to help in the translation of "legacy" data to SGML. Of course, these tools cannot add intelligence to data if it was not there in the legacy format, but they can do a reasonable and lowcost job of converting material for large scale projects where only broad structural information is needed. For those who are familiar with UNIX, the shareware program sgmls and its successor sp are excellent tools for validating SGML documents and can be incorporated in processing programs. There are also ways in which the markup can be minimized. End tags can be omitted in some circumstances, for example in a list where the start of a new list item implies that the previous one has ended.

There is no doubt that SGML is considered expensive by some projects, but the pay-off can be seen many times over further down the line. The quick and dirty solution to a computing problem does not last very long and history has shown how much time can be wasted converting from one system to another or how much data can be lost because it is in a proprietary system. It is rather surprising that the simple notion of encoding what the parts of a document are, rather than what the computer is supposed to do with them, took so long to catch on. Much of the investment in any computer project is in the data and SGML is the best way we know so far of ensuring that the data will last for a long time and that it can be used and re-used for many different purposes. It also ensures that the project is not dependent on one software vendor. Projects are always under pressure to produce results and this can be done simply with SGML documents by using SoftQuad's Panorama SGML viewer.[15] Panorama immediately gives a sense of what is possible and is easy to use.

The amount of encoding is obviously a key factor in the cost and so any discussion about the cost-effectiveness of an SGML project should really always be made with reference to the specific DTD in use and the level of markup to be inserted. (Unfortunately at present this seems to be rarely the case.) It is quite possible, although clearly not sensible, to have a valid SGML document which consists of one start tag at the beginning and one at the end with no other markup in between. At the other extreme each word (or even letter) in the document could have several layers of markup attached to it. What is clear is that the more markup there is, the more useful the document is and the more expensive it is to create. As far as I am aware, little research has been done on the optimum level of markup, but at least with SGML it is possible to add markup to a document later without prejudicing what is already encoded.

SGML does have one fairly significant weakness. It assumes that each document is a single hierarchic structure, but in the real world (at least of the humanities) very few documents are as simple as this.[16] For example, a printed edition of a play has one structure of acts, scenes and speeches and another of pages and line numbers. A new act or scene does not normally start on a new page and so there is no relationship between the pages and the act and scene structure. It is simply an accident of the typography. The problem arises even with paragraphs in prose texts, since a new page does not start with a new paragraph, or a new paragraph with a new page. For well-known editions the page numbers are important, but they cannot easily be encoded in SGML other than as "empty" tags which simply indicate a point in the text, not the beginning and end of a piece of information. The disadvantage here is that the processing of information marked by empty tags cannot make full use of SGML's capabilities. Another example of the same problem is quotations spanning over paragraphs. They have to be closed and then opened again with attributes to indicate that they are really all the same quotation.

For many scholars, SGML is exciting to work with because it opens up so many more possibilities for working with source material. We now have a much better way than ever before of representing in electronic form the kinds of interpretation and discussion which are the basis of scholarship in the humanities. But as we begin to understand this, some new challenges appear.[17] What happens when documents from different sources (and thus different DTDs) are merged into the same database? In theory, computers make it very easy to do this, but how do we merge material that has been encoded according to different theoretical perspectives and retain the identification and individuality of each perspective? It is possible to build some kind of "mega-DTD", but this may become so free in structure that it is difficult to do any useful processing of the material.

Attention must now turn to making SGML work more effectively. Finding better ways of adding markup to documents is a high priority. The tagging could be speeded up by a program which can make intelligent guesses for the tagging based on information it has derived from similar material that has already been tagged, much in the same way as some word class tagging programs "learn" from text that has already been tagged manually. We also need to find ways of linking encoded text to digital images of the same material without the need for hand-coding. Easier ways must be found for handling multiple parallel structures. All research leading to better use of SGML could benefit from a detailed analysis of documents that have already been encoded in SGML. The very fact that they are in SGML makes this easy to do.

NOTES:

[1] Ian Graham's *HTML Sourcebook: a complete guide to HTML 3.0*, 2nd edition, Wiley, 1996, especially the beginning of Chapter 3, gives an excellent overview of the characteristics of a book in the context of a discussion of the design of electronic resources. The third edition of this book was published early in 1997.

[2] Jay David Bolter's *Writing Spaces: the computer, hypertext and the history of writing*, Erlbaum, 1991, expands on some of these ideas. See also George Landow, *Hypertext: the convergence of contemporary critical theory and technology*, Johns Hopkins, 1992, and my own *Knowledge Representation*, a paper commissioned as part of the Getty Art History Information Program (now the Getty Information Institute) Research Agenda for Humanities Computing, published in *Research Agenda for Networked Cultural Heritage*, p. 31-34, Getty Information Institute, and also available at http://www.ahip.getty.edu/agenda/represen.html.

[3] These terms have been used, among others, by the Model Editions Partnership (http://mep.cla.sc.edu).

[4] This was the planning meeting for the Text Encoding Initiative project. It was held in November 1987.

[5] C.J. Date, *An Introduction to Database Systems*, 4th edition, Addison Wesley, 1986 is a good introduction to relational database technology.

[6] By far the most useful starting point for information about SGML is the very comprehensive Web site at http://www.sil.org/sgml/. This is maintained and updated very regularly by Robin Cover of the Summer Institute for Linguistics.

[7] The TEI's Web site is at http://www.uic.edu/orgs/tei/. It contains links to electronic versions of the TEI Guidelines and DTDs as well as projects which are using the DTD.

[8] See Richard Giordano, "The Documentation of Electronic Texts Using Text Encoding Initiative Headers: an Introduction", *Library Resources and Technical Services*, 38 (1994), 389ff for a detailed discussion of the header from the perspective of someone who is both a librarian and a computer scientist.

[9] More information about the EAD can be found at http://lcweb.loc.gov/ead/. This site has examples of the Library of Congress EAD projects. Others can be found via links from the SGML Web site.

[10] This example can be seen at http://www.ceth.rutgers.edu/projects/griffis/project.htm. The site also provides instructions for downloading the Panorama SGML viewer.

[11] See Yuri Rubinsky, "Electronic Texts the Day After Tomorrow", p5-13 in *Visions and Opportunities in Electronic Publishing: Proceedings of the Second Symposium, December 5-8, 1992*, edited by Ann Okerson, Association for Research Libraries, also available at http://arl.cni.org:80/scomm/symp2/Rubinsky.html. Rubinsky was the founder of SoftQuad and a

leading figure in the SGML community until his tragic early death in January 1996.

[12] There is a very useful set of Frequently Asked Questions (FAQ) on XML at http://www.ucc.ie/xml/. See also the XML section of the SGML Web site at http://www.sil.org/sgml/related.html#xml.

[13] See note (3).

[14] The Orlando Project's Web site is at http://www.ualberta.ca/ORLANDO/.

[15] A free version of Panorama can be used as Web helper application. The Professional version runs as a standalone program. It is well within the price range of an academic user and, together with WordPerfect 7, provides a cheap way of beginning to work with SGML.

[16] In order to deal with the problem of overlap, the Wittgenstein Archives at the University of Bergen (http://www.hd.uib.no/wab/) have devised their own encoding scheme MECS (Multi-Element Code System). MECS contains some of the properties of SGML, but has simpler mechanisms for structures which are cumbersome in SGML. However this has meant that they have had to develop their own software to process the material.

[17] For a longer discussion of new questions posed by the use of SGML and especially its perceived lack of semantics, see C.M. Sperberg-McQueen's closing address to the SGML92 conference at http://www.sil.org/sgml/sgml92sp.html. He notes: `In identifying some areas as promising new results, and inviting more work, there is always the danger of shifting from "inviting more work" to "needing more work" and giving the impression of dissatisfaction with the work that has been accomplished. I want to avoid giving that impression, because it is not true, so I want to make very clear: the questions I am posing are not criticisms of SGML. On the contrary, they are its children. .... SGML has created the environment within which these problems can be posed for the first time, and I think part of its accomplishment is that by solving one set of problems, it has exposed a whole new set of problems.'

……

For additional information about the conference, or The Andrew W. Mellon Foundation's scholarly communication initiatives, please contact Richard Ekman. For additional information about ARL or this web site contact Patricia Brennan, ARL Program Officer at (202) 296-2296.

Return to Office of Scholarly Communication Home Page

ARL Home ARL Scholarly Communication and Technology Home Page

© Association of Research Libraries, Washington, DC

Web Design by Angelo F. Cruz

Scholarly Communication and Technology

Conference Organized by The Andrew W. Mellon Foundation

at Emory University
April 24-25, 1997

Copyright © of the papers on this site are held by the individual authors or The Andrew W. Mellon Foundation. Permission is granted to reproduce and distribute copies of these works for nonprofit educational or library purposes, provided that the author, source, and copyright notice are included on each copy. For commercial use, please contact **Richard Ekman** at the The Andrew W. Mellon Foundation.

Session #5 Technical Choices and Standards

Technical Standards and Medieval Manuscripts

Brother Eric Hollas
OSB, Director, Hill Monastic Manuscript Library
St. John's University

• •

Medieval manuscripts, that is, handwritten codices produced between the fifth century and the late fifteenth century, are counted among the greatest intellectual treasures of western civilization. Manuscripts are significant to scholars of medieval culture, to art historians, calligraphers, musicologists, paleographers and other researchers for a multiplicity of reasons. They contain what remains of the classical literary corpus; and they chronicle the development of religion, history, law, philosophy, language and science from the Middle Ages into early modern times.

Even though manuscripts represent the most voluminous surviving artifact from the Middle Ages, the very nature of this resource presents challenges for usage. For one, each manuscript -- as a hand-written document -- is a unique creation. As such, copies of a particular work may contain variances that make all copies -- wherever they might be -- necessary for review by an interested scholar. Secondly, access to unique manuscripts spread across several countries or continents can be both costly and limited. A scholar wishing to consult manuscripts must often travel throughout Europe, the United States and other countries to find and study manuscripts of interest. Such research is costly and time-consuming. The universities, museums and libraries that own these manuscripts may lack the space and personnel to accommodate visiting scholars, and in some cases research appointments need to be arranged months in advance. Compounding these difficulties can be the challenge of inconvenient geography. While eminent collections reside in the great capitals of Europe, other collections of scholarly interest are housed in remote sites with no easy access at all. And finally, the uniqueness of each manuscript presents special issues of preservation. Because manuscripts represent finite and non-renewable resources, librarians concerned with the general wear and tear on manuscripts have begun to restrict access to these codices.

In an effort to preserve medieval manuscripts and to create broader and more economical access to their contents, many libraries have in recent decades sought to provide filmed copies of their manuscripts to users. This has been a long-established practice at such institutions as the British Library, the Bibliotheque National, and the Vatican Library. Additionally, some libraries have been established for the specific purpose of microfilming manuscript collections. The Institut de Recherche et d'Histoire des Textes in Paris, for example, for decades has been filming the manuscripts of the provincial libraries in France. Since its founding in 1965, the Hill Monastic Manuscript Library at Saint John's University in Minnesota has filmed libraries in Austria, Germany, Switzerland, Spain, Portugal, Malta and Ethiopia. And at the Vatican Film Library at Saint Louis University, one can find microfilms of 37,000 manuscript codices from the Biblioteca Apostolica Vaticana in Rome. Instead of traveling from country to country and from library to library, researchers may make a single trip to one of these microfilm libraries to consult texts, or, in certain circumstances, they may order microfilm copy by mail. Microfilm was a great step forward in providing access to manuscripts, and it still offers tremendous advantages of economy and democratic access to scholars. Still, there are certain limitations because in some situations researchers must visit the microfilm institutions to consult directly, and the purchase of microfilm -- even if ordered from a distance -- can entail long waits for delivery. And compounding these difficulties can be the inconsistency or inadequacy of existing descriptions of medieval manuscripts.

Access to manuscripts in particular collections is guided by the finding aids that have been developed through the centuries. The medieval shelf list has given way to the modern catalogue in most cases, but challenges in locating particular manuscripts and in acquiring consistent information abound. Traditionally, libraries in Europe, the United States, and elsewhere have published manuscript catalogues to describe their handwritten books. These catalogues are themselves scholarly works that combine identification of texts with a description of the codex as a physical object. Although these catalogues are tremendously valuable to scholars, they are not without their shortcomings. With respect to manuscript catalogues, there is presently no agreement within the medieval community on the amount and choice of detail reported, on the amount of scholarly discussion provided and on the format of presentation. Moreover, to consult these published books in the aggregate requires access to a research library prepared to maintain an increasingly large collection of expensive and specialized books. And beyond that, the production of a modern catalogue requires expertise of high caliber and the financial resources that facilitate the work. Because many libraries do not have such resources available, many collections have gone uncatalogued or have been catalogued only in an incomplete fashion. The result for the scholar is a paucity of the kind of information that makes manuscript identification and location possible.

Existing and emerging electronic technologies present extraordinary opportunities for overcoming these challenges and underscore the need to create a long-term vision for *Electronic Access to Medieval Manuscripts*. Electronic access both to manuscript images as well as to bibliographic information presents remarkable opportunities. For one, the distance between the manuscript and the reader vanishes -- providing the opportunity for a researcher anywhere to consult the image of a manuscript in even the remotest location. Secondly, electronic access obviates the security issues and the preservation concerns that accompany usage. Furthermore, electronic access will permit the scholar to unite the parts of a manuscript that may have been taken apart, scattered, and subsequently housed at different sites. It also allows for image enhancement and manipulation that conventional reproductions simply do not make available. Electronic access will also make possible comprehensive searches of catalogue records, research information, texts and tools -- with profound implications in terms of cost to the researcher and a more democratic availability of materials to a wider public.

One may imagine a research scenario that contrasts sharply with the conventional methods that have been the mainstay of manuscript researchers. Using a personal computer in an office, home, educational institution or library, scholars will be able to log on to a bibliographic utility (i.e. RLIN or OCLC) or on to an SGML database on the World Wide Web and browse catalogue records from the major manuscript collections around the world. To make this vision a reality requires adherence to standards, however -- content standards to insure that records include the information that scholars need, and encoding standards to insure that that information will be widely accessible both now and in the future.

This point may be demonstrated by considering several computer cataloguing projects developed since the mid-1980's. These efforts include the Benjamin Catalogue for the History of Science, the International Computer Catalog of Medieval Scientific Manuscripts in Munich, the Zentralinventar Mittelalterlicher Handschriften (ZIH) at the Deutsche Staatsbibliothek in Berlin, MEDIUM at the Institut de Recherche et d'Histoire des Textes in Paris and PhiloBiblon at the University of California, Berkeley. The Hill Monastic Manuscript Library has also embarked on several electronic projects to increase and enhance scholarly access to its manuscript resources. In 1985, Thomas Amos, then Cataloguer of Western Manuscripts at HMML, began development of the *Computer Assisted Cataloguing Project*, a relational database which he used to catalogue manuscripts from Portuguese libraries filmed by HMML.

These electronic databases as well as others from manuscript institutions around the world represent an enormous advancement in scholarly communication in the field of manuscript studies. As in the case of printed catalogues and finding aids, however, these data management systems fall short of the ideal on several counts. First, each is a local system that must be consulted on site or purchased independently. Second, the development and maintenance of these various databases involve duplication of time, money and human resources. All rely on locally-developed or proprietary software, and this has posed problems for the long-term maintenance and accessibility of the information. Finally, and probably most importantly, each system contains its own unique set of data elements and rules and procedures for data entry and retrieval. When each of these projects was begun, its founders decided independently what information about a manuscript to record, how to encode it and how to retrieve it. Each of the databases adopted a different solution to the basic problems of description and indexing, and the projects differed from each other with regard to completeness of the data entered and the modes in which it could be retrieved.

The lessons to be drawn from these experiences are clear and enunciate the hazzards for the future if distinctively different approaches are not pursued. First of all, local institutions could not maintain locally developed software and systems. In the instances of projects that chose to rely on proprietary software, it became apparent that the latter was dependent on support from the manufacturer, whose own longevity in business could not be guaranteed, or who could easily abandon such software programs when advances provided new opportunities. Furthermore, experience has demonstrated that it is not always easy to translate such material into other formats, and if modified it poses the same problems of maintenance as locally developed software. Beyond that, different projects made substantially different decisions about record content, and those decisions were sometimes influenced by the software that was available. This lack of consistency made it difficult to disseminate the information gathered by each project, and for their part funding agencies were reluctant to continue their support for such limited projects. All of which reiterates the fundamental need for content standards to insure that records include the information that scholars need and encoding standards to insure the wide accessibility of that information both now and into the future. It is the objective of *Electronic Access to Medieval Manuscripts* to address these issues.

Electronic Access to Medieval Manuscripts is sponsored by the Hill Monastic Manuscript Library, Saint John's University, Collegeville, Minnesota, in association with the Vatican Film Library, Saint Louis University, and has been funded by a grant from The Andrew W. Mellon Foundation. It is a three-year project to develop guidelines for cataloguing medieval and renaissance manuscripts in electronic form. For this purpose it has assembled an international team of experts in manuscript studies and library and information science which will examine the best current manuscript cataloging practice in order to identify the information appropriate to describing and indexing manuscripts on two levels, core and detailed. Core level descriptions, which will contain the basic or minimum elements required for the identification of a manuscript, will be useful for describing manuscripts that have not yet been fully cataloged, and may also be used to give access to detailed descriptions, or to identify the sources of digital images or other information extracted from manuscripts. Guidelines for detailed or full descriptions will be designed to accommodate the kinds of information found in full scholarly manuscript cataloging.

In addition to suggesting guidelines for content, *Electronic Access to Medieval Manuscripts* will also develop standards for encoding both core-level and detailed manuscript descriptions in both MARC and SGML. The MARC (Machine-Readable Cataloging) format underlies most electronic library catalogs in North America and the United Kingdom, and it is used also as a vehicle for international exchange of bibliographic information. MARC bibliographic records are widely accessible through local and national databases, and libraries with MARC-based cataloguing systems can be expected to maintain them for the foreseeable future. SGML (Standardized General Markup Language) is a platform-independent and extremely flexible way of encoding electronic texts for transmission and indexing. It supports the linking of texts and images, and SGML-encoded descriptions are easily converted to HTML for display on the World Wide Web. In developing standards for SGML encoding of manuscript descriptions, *Electronic Access to Medieval Manuscripts* will work closely with the *Digital Scriptorium*, a project sponsored jointly by the Bancroft Library at the University of California, Berkeley, and the Butler Library at Columbia University.

The project working group for *Electronic Access to Medieval Manuscripts* consists of representatives from a number of North American and European institutions. Drafts produced by the working group will be advertised and circulated to the international community of manuscript scholars for review and suggestions. The cataloguing and encoding guidelines that result from the work of the project will be made freely available to any institution that wishes to use them.

For the purposes of *Electronic Access to Medieval Manuscripts*, the standards for cataloguing medieval manuscripts are crucial, but so too is the application of content standards to the two encoding standards whose existence and ubiquitous usage address the issues noted earlier. At the risk of stating the obvious, *Electronic Access to Medieval Manuscripts* has chosen to work with two existing and widely used encoding standards because it is unwise for medievalists to reinvent the wheel and waste resources on solutions that are temporary and which will require added resources to take them into future applications.

With regard to encoding standards, the universal acceptance of MARC and the accessibility of MARC records on-line make it a particularly attractive option. But there are other compelling reasons that make MARC an excellent choice. First, most libraries already have access to a bibliographic utility (such as OCLC and RLIN) that utilizes MARC-based records, and these institutions have invested considerable resources in creating catalogue records for their printed books and other collections. Second, since most catalogue records for printed books and reference materials are already in MARC-based systems, placing manuscript records in the same system makes good sense from the standpoint of proximity and one-stop searching. Third, by using MARC, local libraries need not develop or maintain their own database systems. Finally, although it may be unrealistic to expect that all manuscript catalogue records will one day reside in a single database, therefore allowing for a universal search of manuscript records, it is far more likely that a majority of manuscript institutions in the United States will be willing to place their manuscript records in this bibliographic utility rather than in other existing environments. Thus the value of selecting MARC as an encoding standard seems clear. MARC systems exist; they are widely accessible; they are supported by other broader interests; and enough bibliographic data already exists in MARC to guarantee its maintenance or its automatic transfer to any future platform. In USMARC (RLIN and OCLC databases) there are already a significant number of records for medieval manuscripts or microfilms of them, prepared and entered by the various institutions that hold these items. Regrettably, there is generally little consistency in description, indexing or retrieval for these records; all of which points back to the need for standards for content as well as encoding standards. Furthermore, MARC as it currently exists has limits in its abilities to describe medieval manuscripts (e.g.: it does not provide for the inclusion of *incipits*), but nonetheless it offers possibilities for short records that point to broader sets of data in other contexts. Still, MARC, with its records in existing bibliographic databases, is particularly advantageous for small institutions with few manuscript holdings, and it remains for them perhaps the most promising vehicle for disseminating information about their collections.

The second viable encoding option, particularly in light of the recent success of the Archival Finding Aid Project at the University of California, Berkeley, is the use of Standard Generalized Markup Language (SGML). As a universal standard for encoding text, SGML can be used to encode and index catalogue records and other data including text, graphics, images and multimedia objects such as video and sound. A more flexible tool than MARC, SGML is more easily adapted to complex hierarchical structures such as traditional descriptions of medieval manuscripts, and it offers broad possibilities for encoding and indexing existing, as well as new, manuscript catalogues. As an encoding scheme, SGML demonstrates its value as a non-proprietary standard. In many respects it is much more flexible than MARC or any established database program, and it is possible to write a Document Type Definition (DTD) taking into account the particular characteristics of any class of document. SGML offers the further advantage that encoded descriptions can be linked directly to digital images, sound clips (e.g., for musical performances) or other bodies of digital information relating to a manuscript. Numerous initiatives using SGML suggest great promise for the future. The experience of the American archival profession with the Encoded Archival Description (EAD) suggests that this can be a good approach to encoding manuscript descriptions, which have many structural analogies to archival finding aids. The Canterbury Tales project, based at Oxford, has demonstrated that SGML, based on a Text Encoding Initiative (TEI) format, can be used successfully to give sophisticated access to images of manuscripts, text transcriptions and related materials. In addition, several English libraries have already experimented with SGML DTD's, mostly TEI-conformant, for manuscripts. And finally, MASTER, an Oxford-based group, is interested in developing a standard DTD for catalogue descriptions of medieval manuscripts, and it and *Electronic Access to Medieval Manuscripts* have begun to coordinate their efforts toward achieving this common goal.

The emerging interconnectivity of MARC and SGML presents tremendous opportunities for *Electronic Access to Medieval Manuscripts*. Currently there is work on a DTD for the MARC format that will allow automatic conversion of MARC encoded records into SGML. Recently, a new field (856) was added to the MARC record that will accommodate web addresses. Implementation of this field will allow researchers seeking access to a cataloguing record in a bibliographic utility to read the URL (Universal Resource Locator) and then enter the address into a web browser and link directly to a website containing a detailed manuscript record or other scholarly information. In the future, for researchers who enter the bibliographic utility through a web browser, this will be an active hypertext link. *Electronic Access to Medieval Manuscripts* envisions an environment in which institutions can enter their manuscript catalogue records into MARC, display them in a bibliographic utility to maximize economy and access, and then embed a hypertext link to a more detailed catalogue record, an image file or scholarly information on an SGML server.

It has been the cumulative experience of recent years that has shaped the development and goals of *Electronic Access to Medieval Manuscripts*. Concerned with arriving at standards for cataloguing manuscripts in an electronic environment, the project seeks to provide standards for both core and full or detailed level manuscript records that will serve the expectations and needs of scholars who seek consistent information from one library to another, while they will afford flexibility to those cataloguers and libraries wishing to provide various levels of information about their individual manuscripts. In structuring its program and goals, *Electronic Access to Medieval Manuscripts* also has sought to arrive at guidelines for encoding into MARC and SGML formats that will provide useful, economic and practical long-term alternatives to the libraries which select one of these options in the future.

....

For additional information about the conference, or The Andrew W. Mellon Foundation's scholarly communication initiatives, please contact Richard Ekman. For additional information about ARL or this web site contact Patricia Brennan, ARL Program Officer at (202) 296-2296.

Return to Office of Scholarly Communication Home Page

ARL Home ARL Scholarly Communication and Technology Home Page

© Association of Research Libraries, Washington, DC

Web Design by Angelo F. Cruz
Maintained by ARL Web Administrator
Last Modified: August 1, 1997

Scholarly Communication and Technology

Conference Organized by The Andrew W. Mellon Foundation

at Emory University
April 24-25, 1997

Copyright © of the papers on this site are held by the individual authors or The Andrew W. Mellon Foundation. Permission is granted to reproduce and distribute copies of these works for nonprofit educational or library purposes, provided that the author, source, and copyright notice are included on each copy. For commercial use, please contact Richard Ekman at the The Andrew W. Mellon Foundation.

Session #5 Technical Choices and Standards

Digital Image Quality: From Conversion to Presentation and Beyond

Anne R. Kenney
Associate Director, Department of Preservation
Cornell University Library

. . . .

As this conference attests, there are a number of significant digital library projects underway designed to test the economic value of digital over physical library building. Business cases are being developed to demonstrate the economics of digital applications to assist research and cultural institutions respond to the challenges of the information explosion, spiraling storage and subscription costs, and increasing user demands. These projects also reveal that the costs of selecting, converting, and making digital information available can be staggering, and that the costs of archiving and migrating that information over time are not insignificant.

Economic models comparing the digital to the traditional library show that digital will become more cost-effective *provided* the following four assumptions prove true:

- that digital collections can alleviate the need to support full traditional libraries at the local level,

- that use will increase with electronic access, and
- that the long-term value of digital collections will exceed the costs associated with their creation, maintenance, and delivery.[1]

These four assumptions--resource sharing, lower costs, meeting user demands for timely and enhanced access, and continuing value of information--presume that electronic files will have relevant content and meet baseline measures of functionality over time. Although a number of conferences and publications have addressed the need to develop selection criteria for digital conversion, and to evaluate the effective use of digitized material, more rhetoric than substantive information has emerged regarding the impact on scholarly research of creating digital collections and making them accessible over networks.

As has been argued elsewhere, I believe that digital conversion efforts will prove economically viable only if they focus on creating electronic resources for long-term use. Retrospective sources should be selected carefully based on their intellectual content; digital surrogates should effectively capture that intellectual content; and access should be more timely, usable, or cost-effective than is possible with original source documents. In sum, I would argue that long-term utility should be defined by the informational value and functionality of digital images, not limited by technical decisions made at the point of conversion or anywhere else along the digitization chain. In this paper, I advocate a strategy of "full informational capture" to ensure that digital objects rich enough to be useful over time are created in the most cost-effective manner.[2]

There is much to be said for capturing the best possible digital image you can. From a preservation perspective, the advantages are obvious. An "archival" digital master can be created to replace rapidly deteriorating originals or to reduce storage costs and increase access times to office back files, provided the digital surrogate is a trusted representation of the hardcopy source. It also makes economic sense, as Michael Lesk has noted, to "turn the pages once" and produce a sufficiently high level image so as to avoid the expense of reconverting at a later date when technological advances require, or can effectively utilize, a richer digital file.[3] This economic justification is particularly compelling as the labor costs associated with identifying, preparing, inspecting, and indexing digital information far exceed the costs of the scan itself. In recent years, the costs of scanning and storage have declined rapidly, narrowing the gap between high quality and low quality digital image capture. Once created, the archival master can then be used to create derivatives to meet a variety of current and future user's needs: high resolution may be required for printed facsimiles, on-screen detailed study, [4] and in the future for intensive image processing; moderate to high resolution for character recognition systems and image summarization techniques;[5] and lower resolution images, encoded text, or PDFs derived from the digital masters for on-screen display and browsing.[6] The quality, utility, and expense of all these derivatives will be directly affected by the quality of the initial scan.[7]

If there are compelling reasons for creating the best possible image, there is also much to be said for not capturing more than you need. At some point, adding more resolution will not result in greater quality, just a larger file size and higher costs. The key is to match the conversion process to the informational content of the original. At Cornell, we've been investigating digital imaging in a preservation context for eight years. For the first three years, we concentrated on what was technologically possible--on determining the best image capture we could secure. For the last five years, we've been striving to define the minimal requirements for satisfying informational capture needs. No more, no less.

Digital Benchmarking

To help us determine what is minimally acceptable, we have been developing a methodology, called benchmarking. Digital benchmarking is a systematic procedure to forecast a likely outcome. It begins with a assessment of the source documents and user needs; factors in relevant objective and subjective variables associated with stated quality, cost, and/or performance objectives; involves the use of formulas that represent the inter-relationship of those variables to desired outcomes; and concludes with confirmation through carefully structured testing and evaluation. If the benchmarking formula does not consistently predict the outcome, it may not contain the relevant variables or reflect their proper relationship--and it should be revised.

Benchmarking does not provide easy answers, but a means against which to evaluate possible answers for how best to balance quality, costs, timeliness, user requirements, and technological capabilities in the conversion, delivery, and maintenance of digital resources. It is also intended as a means to formulate a range of possible solutions on the macro level rather than on an individual, case-by-case basis. For many aspects of digital imaging, benchmarking is still unchartered territory. Much work remains to be able to define conversion requirements for certain document types, e.g., photographs and high end book illustrations; for conveying color information; for evaluating the effects of new compression algorithms; and for providing access on a mass scale to a digital database of material representing a wide range of document types and document characteristics. .

We began benchmarking with the conversion of printed text. We anticipate that within 2 years, quality benchmarks for image capture and presentation of the broad range of paper and film based research materials--including manuscripts, graphic art, halftones, and photographs--will be well defined through a number of projects currently underway.[8] In general, these projects are designed to be system independent and are based increasingly on assessing the attributes and functionality characteristic of the source documents themselves, coupled with an understanding of user perceptions and requirements.

Why benchmarking?

Because there are no standards for image quality, because different document types require different scanning processes, there is no "silver bullet" for conversion. This frustrates many librarians and archivists who are seeking a simple solution to a complex issue. I suppose if there really were the need for a silver bullet, I'd recommend that most source documents be scanned at a minimum of 600 dpi with 24 bit color, but that would result in tremendously large file sizes, and a hefty conversion cost. One would also be left with the problems of transmitting and displaying those images.

We began benchmarking with conversion, but we are now applying this approach to the presentation of information on screen. The number of variables that govern display are many, and it will come as no surprise that they preclude the establishment of a single best method for presenting digital images. But here, too, the urge is strong to seek a single solution. If display requirements paralleled conversion requirements--that is, if a 600 dpi, 24 bit image had to be presented on screen, then *at best*, with the highest resolution monitors commerically available, only documents whose physical dimensions did not exceed 2.7" x 2.13" could be displayed--and they could not be displayed at their native size. Now most of us are interested in converting and displaying items that are larger than postage stamps, so these "simple solutions" are for most purposes impractical, and compromises will have to be made.

The object of benchmarking is to make informed decisions about a range of choices and to understand in advance the consequences of such decisions. The benchmarking approach can be applied across the full continuum of the digitization chain, from conversion to storage to access to presentation. Our belief at Cornell is that benchmarking must be approached holistically, that it is essential to understand at the point of selection what the consequences downstream for conversion and presentation will be. This is especially important as institutions consider inaugurating large scale conversion projects. Towards this end, the advantages of benchmarking are several in number.

1. Benchmarking is first and foremost a management tool, designed to lead to informed decision-making. It offers a starting point and a means for narrowing the range of choices to a manageable number. Although clearly benchmarking decisions must be judged through actual implementations, the time spent in experimentation can be reduced, the temptaton to overstate or understate requirements may be avoided, and the initial assessment requires no specialized equipment nor expenditure of funds. Benchmarking allows one to scale knowledgeably, to make decisions on a macro level, rather than to determine those requirements through item-by-item review or by setting requirements for groups of materials that may be adequate for only a portion of them.

2. Benchmarking provides a means for interpreting vendor claims. If you have spent any time reading product literature, you may have become convinced, as I have, that the sole aim of any company is to sell its product. Technical information will be presented in the most favorable light, which is often incomplete and intended to discourage product comparisons. One film scanner for instance may be advertised as having a resolution of 7500 dpi; another may claim 400 dpi. In fact, these two scanners could provide the very same capabilities but it may be difficult to reach that conclusion without additional information. You may end up spending considerable time on the phone, first getting past the marketing representatives, and then questioning closely those with a technical understanding of the product's capabilities. If you have benchmarked your requirements, you will be able to focus the discussion on your particular needs.

3. Benchmarking can assist you in negotiating with vendors for services and products. I've spent many years advocating the use of 600 dpi bitonal scanning for printed text and invariably when I begin a discussion with a representative of an imaging service bureau, he will try to talk me out of that high a resolution, claiming that I do not need it or that it will be exhorbitantly expensive. I suspect he is in part motivated to make those claims because he believes them, and in part because his company may not provide that service and he wants my business. If I had not benchmarked my resolution requirements, I might be pursuaded by what this salesperson has to say.

4. Benchmarking can lead to careful management of resources. If you know up front what your requirements are likely to be and the consequences of those requirements, you can develop a

budget that reflects the actual costs, identify prerequisites for meeting those needs, and, perhaps most important, avoid costly mistakes. Nothing will doom an imaging project more quickly than buying the wrong equipment or having to manage image files that are not supported by your institution's technical infrastructure.

5. Benchmarking can also allow you to predict what you can deliver under specific conditions. It is important to understand that an imaging project may break at the weakest link in the digitization chain. For instance, if your institution is considering scanning its map collection, one should be realistic about what ultimately can be delivered to the user at her desktop. Benchmarking lets you predict how much of the image and what level of detail contained therein can be presented on-screen for various monitors. Even with the most expensive monitor available today, presenting oversize material completely, with small detail intact, is impractical.

How Does It Work?

Having spent some time extolling the virtues of digital benchmarking, I'd like to turn next to describing this methodology as it applies to conversion, and then to move to a discussion of on-screen presentation.

Objective Evaluation:

Determining what constitutes informational content becomes the first step in the conversion benchmarking process. This can be done objectively or subjectively. Let's consider an objective approach first. One way to do this would be to peg conversion requirements to the process used to create the original document. Take resolution, for instance. Film resolution can be measured by the size of the silver grains suspended in an emulsion, whose distinct characteristics are appreciated only under microscopic examination. Should we aim for capturing the properties of the chemical process used to create the original? Or should we peg resolution requirements at the recording capability of the camera or printer used?

There are objective scientific tests that can measure the overall information carrying capacity of an imaging system, such as the Modulation Transfer Function, but such tests require expensive equipment and are still beyond the capabilities of most outside industry or research labs. In practical applications, the resolving power of a microfilm camera is measured by means of a technical test chart where the distinct number of black and white lines discerned is multiplied by the reduction ratio used to determine the number of line pairs per millimeter. A system resolution of 120 line pairs per millimeter is considered good; above 120 is considered excellent. To capture digitally all the information present on a 35mm frame of film with a resolution of 120 lppm would take a bitonal film scanner with a pixel array of 12,240.[9] There is no such beast on the market today.

How far down this path should we go? It may be appropriate to require that the digital image accurately depict the gouges of a wood cut or the scoops of a stipple engraving, but what about the exact dot pattern and screen ruling of a halftone? the strokes and acid bite of an etching? the black lace of an aquatint that only becomes visible at a magnification above 25x? Offset publications are printed at 1200 dpi--should we chose that resolution as our starting point for scanning text? Significant information may well be present at that level in some cases, as may be argued for medical x-rays, but in other cases, attempting to capture all possible information will far exceed the inherent properties of the image as distinct from the medium and process used to create it. Consider for instance a 4 *x* 5 negative of a badly blurred photograph. The negative is incredibly information dense, but the information it conveys is not significant.

Obviously, any practical application of digital conversion would be overwhelmed by the recording, computing, and storage requirements that would be needed to support capture at the structure or process level. Although offset printing may be produced at 1200 dpi, most individuals would not be able to discern the difference between a 600 dpi and a 1,000 dpi digital image of that page, even under magnification. In choosing the higher resolution one would be adding more bits, increasing the file size, but with little to no appreciable gain. The difference between 300 dpi and 600 dpi, however, can be easily observed, and, in my opinion, is worth the extra time and expense to obtain. The relationship between resolution and image quality is not linear: at some point as resolution increases, the gain in image quality will level off. Benchmarking will help you to determine where the leveling begins.

Subjective Evaluation:
I would argue, then, that determining what constitutes informational content is best done subjectively. It should be based on an assessment of the attributes of the document rather than the process used to create that document. Reformatting via digital--or analog--techniques presumes that the essential meaning of an original can somehow be captured and presented in another format. There is always some loss of information when an object is copied. The key is to determine whether that informational loss is significant or not. Obviously for some items, particularly those of intrinsic value, a copy can only serve as a surrogate, not as a replacement. This determination should be made by those with curatorial responsibility and a good understanding of the nature and signficance of the material. Those with a trained eye should consider the attributes of the document itself as well as the immediate and potential uses that researchers will make of its informational content.

Determining Scanning Resolution Requirements For Replacement Purposes:

To illustrate benchmarking for conversion, let's consider the brittle book. For brittle books published during the last century and a half, detail has come to represent the size of the smallest significant character in the text, usually the lower case "e." To capture this information--which consists of black ink on a light background--resolution is the key determinant of image quality.

Benchmarking resolution requirements in a digital world has its roots in micrographics, where standards for predicting image quality are based on the Quality Index (QI). QI provides a means for relating system resolution and text legibility. It is based on multiplying the height of the smallest significant character "h" by the smallest line pair pattern resolved by a camera on a technical test target, "p," QI=h x p. The resulting number is called the Quality Index, and it is used to forecast levels of image quality--marginal (3.6), medium (5.0) or high (8.0)--that will be achieved on the film. This approach can be used in the digital world, but a number of adjustments must be made to account for the differences in the ways in which microfilm cameras and scanners capture detail.[10] Specifically, it is necessary to:

1. Establish levels of image quality for digitally rendered characters that are analogous to those established for microfilming (illustration showing differences in quality degradation). Note that in photographically reproduced images, quality degradation results in a fuzzy or blurred image. Usually degradation with digital conversion is revealed in the ragged or stairstepped appearance of diagonal lines or curves, known as aliasing or "jaggies."

2. Rationalize system measurements. Digital resolution is measured in dots per inch; classic resolution is measured in line pairs per millimeter. To calculate QI based on scanning resolution, one must convert from one to the other. One millimeter equals .039 inches, so to determine the number of dots per millimeter, you will need to multiply the DPI by .039.

3. Equate dots to line pairs. Again, classic resolution refers to line *pairs* per millimeter (one black line and one white line), and since a dot occupies the same space as a line, two dots must be used to represent one line pair. This means the dpi must be divided by two to be made equivalent to "p."

With these adjustments, we can modify the QI formula to create a digital equivalent. From QI= p*x* h,

we now have $QI = \frac{.039dpi}{2} x\ h$

which can be simplified to .0195dpi x h.

For bitonal scanning, we would also want to adjust for possible misregistration due to sampling errors brought about in the thresholding process in which all pixels are reduced to either black or white. To be on the conservative side, the authors of AIIM TR26-1993 advise increasing the input scanning resolution by at least 50% to compensate for possible image detector mis-alignment. The formula would then be

$QI = \frac{.039dpi}{3} x\ h$ which can be simplified to .013dpi x h.

So how does all this work?

Consider a printed page that contains characters measuring 2mm high and above. If the page were scanned at 300 dpi, what level of quality would you expect to obtain? By plugging in the dpi and the character height and solving for QI, you would discover that you can expect a QI of 8, or excellent rendering.

One can also solve the equation for the other variables. Consider for example a scanner with a maximum of 400 dpi. You can benchmark the size of the smallest character that you could capture with medium quality (a QI of 5), which would be .96mm high. Or you can calculate the input scanning resolution required to achieve excellent rendering of a character that is 3 mm high (200 dpi).

With this formula, and an understanding of the nature of your source documents, you can benchmark the scanning resolution needs for printed material. We took this knowledge and applied it to the types of documents we were scanning--brittle books published from 1850-1950. We reviewed printers' type sizes commonly used by publishers during this period, and discovered that virtually none utilized type fonts smaller than 1 mm in height, which, according to our benchmarking formula, could be captured with excellent quality using 600 dpi bitonal scanning. We then tested these benchmarks by conducting an extensive on-screen and in print examination of digital facsimiles for the smallest font-sized Roman and non-Roman type scripts used during this period. This verification process confirmed that an input scanning resolution of 600 dpi was indeed sufficient to capture the monochrome text-based information contained in virtually all books published during the period of paper's greatest brittleness. Although many of those books do not contain text that is as small as 1 mm in height, a sufficient number of them do. To avoid the labor and expense of performing item by item review, we currently scan all books at 600 dpi resolution.[11]

Conversion Benchmarking Beyond Text

Although we've conducted most of our experiments on printed text, we are beginning to benchmark resolution requirements for non-textual documents as well. For non-text based material, we have begun to develop a benchmarking formula that would be based on the width of the smallest stroke or mark on the page rather than a complete detail. This approach was used by the Nordic Digital Research Institute to determine resolution requirements for the conversion of historic Icelandic maps, and is being followed in the current New York State Kodak Photo CD project being conducted at Cornell on behalf of the Eleven Comprehensive Research Libraries of New York State. The measurement of such fine detail will require the use of a 25-50x lupe with a metric hairline that differentiates below .1mm.

Benchmarking for conversion can be extended beyond resolution to tonal reproduction (both grayscale and color), to the capture of depth, overlay, and translucency, to assessing the effects of compression techniques and levels of compression used on image quality, to evaluating the capabilities of a particular scanning methodology, such as the Kodak Photo CD format. It can also be used for evaluating quality requirements for a particular category of materials, e.g., halftones, or to examine the relationship between the size of the document and the size of its significant details, a very challenging relationship which affects both the conversion and the presentation of maps, newspapers, architectural drawings, and other oversized, highly detailed source documents.

Benchmarking involves both subjective and objective components. There must be the means to establish levels of quality (through technical targets, samples of acceptable materials), the means to identify and measure significant information present in the document, the means to relate one to another via a formula, and the means to judge results on-screen and in print for a sample group of documents. Armed with this information, benchmarking enables informed decision making--which often leads to a balancing act involving tradeoffs between quality and cost, between quality and completeness, between completeness and size, or quality and speed.

Benchmarking Display Requirements:

Quality assessments can be extended beyond capture requirements to the presentation and timeliness of delivery options. We begin our benchmarking for conversion with the attributes of the *source documents*. We begin our benchmarking for display with the attributes of the *digital images*.

I believe that all researchers in their heart of hearts expect three things from displayed digital images: they want the full size image to be presented on screen; they expect legibility and adequate color rendering, and they want images to be displayed quickly. Of course they want lots of other things, too, such as the means to manipulate, annotate, and compare images, and for text-based material, they want to be able to conduct key word searches across the images. But for the moment, let's just consider those three requirements: full image, full detail and tonal reproduction, quick display.

Unfortunately, for many categories of documents, satisfying all three criteria at once will be a problem, given the limitations of screen design, computing capabilities, and network speeds. Benchmarking screen display must take all these variables into consideration *and* the attributes of the digital images themselves as user expectations are weighed one against the other. We are just beginning to investigate this interrelationship at Cornell, and although our findings are still tentative and not broadly confirmed through experimentation, I'm convinced that display benchmarking will offer the same advantages as conversion benchmarking to research institutions that are beginning to make their materials available electronically.[12]

Now for the good news: it is easy to display the complete image and it is possible to display it quickly. It is easy to ensure screen legibility--in fact intensive scrutiny of highly detailed information is facilitated on screen. Color fidelity is a little more difficult to deliver, but progress is occurring on that front.[13]

Now for the not so good news: given common desktop computer configurations, it may not be possible to deliver full 24-bit color to the screen--the monitor may have the native capability but not enough video memory or its refresh rate can not sustain a non-flickering image. The complete image that is quickly displayed may not be legible. A highly detailed image may take a long time to deliver and only a small percent of it will be seen at any given time. You may call up a photograph of Yul Brenner only to discover you have landed somewhere on his bald pate.

Benchmarking will allow you to predict in advance the pros and cons of digital image display. Conflicts between legibility and completeness, between timeliness and detail, can be identified and compromises developed. Benchmarking allows you to predetermine a set process for delivering images of uniform size and content, and to assess how well that process will accommodate other document types. Scaling to 72 dpi and adding 3 bits of gray may be a good choice for technical reports produced at 10 point type and above, but will be totally inadequate for delivering digital renderings of full-size newspapers.

To illustrate benchmarking as it applies to display, consider the first two user expectations: complete display and legibility. We expect printed facsimiles produced from digital images to look very similar to the original. They should be the same size, preserve the layout, and convey detail and tonal information that is faithful to the original. Many readers assume that the digital image on screen can also be the same, that if the page were correctly converted, it could be brought up at approximately the same size and with the same level of detail as the original. It is certainly possible to scale the image to be the same size as the original document, but chances are information contained therein will not be legible.

If the scanned image's dpi does not equal the screen dpi, then the image on-screen will either appear larger or smaller than the original document's size. Because scanning dpi most often exceeds the screen dpi, the image will appear larger on the screen--and chances are not all of it will be represented at once. This is because monitors have a limited number of pixels that can be displayed both horizontally and vertically. If the number of pixels in the image exceed those of the screen and the scanning dpi is higher, the image will be enlarged on the screen and not completely presented.

The problems of presenting completeness, detail, and native size are more pronounced in display than in printing. In the latter, industry is capable of very high printing resolutions, and the total number of dots that can be laid down for a given image is great, enabling the creation of facsimiles that are the same size--and often with the same detail--as the original.

The limited pixel dimensions and dpi of monitors can be both a strength and a weakness. On the plus side, detail can be presented more legibly and without the aid of a microscope, which for those conducting extensive textual analysis may represent a major improvement over reviewing the source documents themselves. For instance, papyrologists can rely on monitors to provide the enlarged view of fragment details required in their study. When the originals themselves are examined, they are typically viewed under a microscope at 4 to 10x magnification.[14] Art historians can zoom in on high resolution images to enlarge details or to examine brush strokes that convey different surfaces and materials.[15] On the down side, because the screen dpi is often exceeded by the scanning dpi, and screens have very limited pixel dimensions, many documents can not be fully displayed *if* legibility must be conveyed. This conflict between overall size and level of detail is most apparent when dealing with oversized material, but it also affects a surprisingly large percentage of normal-sized documents as well.

Consider the physical limitations of computer monitors:

Typical monitors offer resolutions from 640 x 480 at the low end to 1600 x 1200 at the high end. The lowest level SVGA monitor offers the possibility of displaying material at 1024 x 768. These numbers, known as the pixel matrix, refer to the number of horizontal by vertical pixels painted on the screen when an image appears.

In product literature, monitor resolutions are often given in dpi which can range from 60 to 120, depending on the screen width and horizontal pixel dimension. The screen dpi can be a misleading representation of a monitor's quality and performance. For example, when SVGA resolution is used on a 14", 17", and 21" monitor, the screen dpi decreases as screen size increases. We might intuitively expect image resolution to increase with the size of the monitor, not decrease. In reality the same amount of an image--and level of detail--would be displayed on all three monitors set to the same pixel dimensions. The only difference would be that the image displayed on the 21 inch monitor would appear enlarged compared to the same image displayed on the 17 and 14 inch monitors.

The pixel matrix of a monitor limits the number of pixels of a digital image that can be displayed at any one time. And, if there is insufficient video memory, you will also be limited to how much gray or color information can be supported at any pixel dimension. For instance, while the three-year old 14" SVGA monitor on my desk supports a 1024 x 768 display resolution, it came bundled with half a megabyte of video memory. It can not display an 8-bit grayscale image at that resolution and it can not display a 24 bit color image at all, even if it is set at the lowest resolution of 640 x 480. Even if I increased its VRAM, I would be bothered by an annoying flicker, as the monitor's refresh rate is not great enough to support a stable image on screen at higher resolutions. It is not coincidental that while the most basic SVGA monitors can support a pixel matrix of 1024 x 768, most of them come packaged with the monitor set at a resolution of 800 x 600. As others have noted, network speeds and the limitations of graphical user interfaces will also affect profoundly user satisfaction with on-screen presentation of digital images.

So how does benchmarking for display work?

Consider the brittle book and how best to display it. Recall that it may contain font sizes at 1 mm and above, so we have scanned each page at 600 d pi, bitonal mode. Let's assume that the typical page averages 4" x 6" in size. The pixel matrix of this image will be: 4 x 600 by 6 x 600, or 2400 x 3600--far above any monitor pixel matrix currently available. Now if I want to display that image at its full scanning resolution on my monitor, set to the default resolution of 800 x 600, it should be obvious to many of you that I will be showing only a small portion of that image--approximately 5% of it will appear on the screen. Let's suppose I went out and purchased a $2,500 monitor that offered a resolution of 1600 x 1200. I'd still only be able to display less than a fourth of that image at any one time.

Obviously for most access purposes, this display would be unacceptable. It requires too much scrolling or zooming out to study the image. If it is an absolute requirement that the full image be displayed with all details fully rendered, I'd suggest converting only items whose smallest significant detail represents nothing smaller than one third of 1% of the total document surface. This means that if you had a document with a one millimeter high character that was scanned at 600 dpi and you wanted to display the full document at its scanning resolution on a 1024 x 768 monitor, the document's physical dimensions could not exceed 1.7" (horizontal) x 1.3" (vertical). This may work well for items such as papyri which are relatively small, at least as they have survived to the present. It also works well for items that are physically large *and* contain large-sized features, such as posters that are meant to be viewed from a distance. If the smallest detail on the poster measured one inch, the poster could be as large as 42" x 32" and still be fully displayed with all detail intact.[16]

Most images will have to be scaled down from their scanning resolutions for on screen access, and this can occur a number of ways. Let's first consider full display on the monitor, and then consider legibility. In order to display the full image on a given monitor, the image pixel matrix must be reduced to fit within the monitor's pixel dimensions. The image is scaled by setting one of its pixel matrixes to the corresponding pixel dimension of the monitor.[17]

To fit the complete page image from our brittle book on a monitor set at 800 *x* 600, we would scale the vertical dimension of our image to 600; the horizontal dimension would be 400 to preserve the aspect ratio of the original. By reducing the 2400 x 3600 pixel image to 400 x 600, we will have discarded 97% of the information in the original. The advantages to doing this are several: it facilitates browsing by displaying the full image, it decreases file size which in turn decreases the transmission time. The down side should also be obvious. There will be a major decrease in image quality as a significant number of pixels are discarded. In other words, the image can be fully displayed, but the information contained in that image may not be legible. To determine whether that information will be useful, we can turn to the use of benchmarking formulas for legible display:

Benchmarking resolution formulas for scaling bitonal and grayscale images for on-screen display:[18]

$\text{dpi} = QI/(.03h)$

$QI = \text{dpi} \, x .03h$

$h = QI/(.03\text{dpi})$

Note: Recall that in the benchmarking resolution formulas for conversion, dpi refers to the scanning resolution. In the scaling formulas, dpi refers to the image dpi (not to be confused with the monitor's dpi).

Let's return to the example of our 4 x 6 brittle page.

If we assume we need to be able to read the 1 mm high character, but that it doesn't have to be fully rendered, then we set our QI requirement at 3.6, which should ensure legibility of characters in context. We can use the benchmarking formula to predict the scaled image dpi:

dpi = QI/.03h, or

dpi = 3.6/(.03x 1), or

dpi =120

The image could be fully displayed with minimal legibility on a 120 dpi monitor. The pixel dimensions for the scaled image would be 120 *x* 4 by 120 *x* 6, or 480 *x* 720. This full image could be viewed on SVGA monitors set at 1024 x 768 or above; slightly over 80% of it could be viewed on my monitor set at 800 x 600.

We can also use this formula to determine a preset scaling dpi for a group of documents to be conveyed to a particular clientele. Consider a scenario where your primary users have access to monitors that can support effectively an 800 x 600 resolution. We could decide whether the user population would be satisfied with receiving only 80% of the document if it meant that they could read the smallest type, which may occur only in footnotes. If your users are more interested in quick browsing, you might want to benchmark against the body of the text, rather than the smallest typed character. For instance, if the main text were in 12 point type and the smallest "e" measured 1.6 mm in height, then our sample page could be sent to the screen with a QI of 3.6 at a pixel dimension of 300 *x* 450, or an image dpi of 75--well within the capabilities of the 800 x 600 monitor.

One can also benchmark the time it will take to deliver this image to the screen--if your clientele are connected via ethernet, this image (with 3 bits of gray added to smooth out rough edges of characters and improve legibility) could be sent to the desktop in under a second--providing readers with full display of the document, legibility of the main text, and a timely delivery. If your readers are connected to the ethernet via a 9600 baud modem, however, the image will take 42 seconds to be delivered. If the footnotes must be readable, the full text can not be delivered at once and the time it will take to retrieve the image will increase. Benchmarking allows you to identify these variables and consider the tradeoffs/compromises associated with optimizing any one of them.

Conclusion:

Benchmarking is an approach, not a prescription. It offers a means to evaluate choices for how best to balance quality, costs, timeliness, user requirements, and technological capabilities in the conversion, delivery, and presentation of digital resources. The value of this approach will best be determined by extensive field testing. We at Cornell are committed to further refinement of the benchmarking methodology, and urge others to consider its utility *before* they commit considerable resources to bringing about the brave new world of digitized information.

FOOTNOTES:

[1] Stephen Chapman and Anne R. Kenney, "Digital Conversion of Library Research Materials: A Case for Full Informational Capture," *D-Lib Magazine,* October 1996.

[2] Currently, scanning is the most cost-effective means to create digital files, and digital imaging is the only electronic format that can accurately render the information, page layout, and presentation of source documents, including text, graphics, and evidence of age and use. By producing digital images, one can create an authentic representation of the original at minimal cost, and then derive the most useful version and format (e.g., marked-up text) for transmission and use.

[3] Michael Lesk, *Image Formats for Preservation and Access. A Report of the Technology Assessment Advisory Committee to the Commission on Preservation and Access,* July 1990; see also Lesk, *Substituting Images for Books: The Economics for Libraries*, January 1996.

[4] See, Charles S. Rhyne, *Computer Images for Research, Teaching, and Publication in Art History and Related Disciplines*, Commission on Preservation and Access, January 1996, p. 4, where he argues that "with each jump in [on-screen image] quality, new uses become possible."

[5] Interesting work is being conducted at Xerox PARC on image summarization, see Francine R. Chen and Dan S. Bloomberg, "Extraction of Thematically Relevant Text from Images," to appear in SDAIR,'96, pp. 163-178.

[6] An interesting conclusion from a project on the use of art and architectural images at Cornell focused on image size guidelines to support a range of user activities. For browsing, the project staff found that images must be large enough for the user to identify the image, but small enough to allow numerous images to be viewed simultaneously--the physical size on the screen preferred by users was 1.25 to 2.25 inches square. For view images in their entirety, images were sized to fit within a 5.5 inch square; for studying, detailed views covering the entire screen were necessary, and for "authoring" presentations or other multimedia projects, users preferred images that fit in a half inch square. See Noni Korf Vidal, Thomas Hickerson, and Geri Gay, "Developing Multimedia Collection and Access Tools, Appendix V. Guidelines for the Display of Images." pp. 14-17. April 1996.

[7] A number of leading experts advocate this approach, including Michael Ester of Luna Imaging, Inc. See for example: Ester, Michael, "Digital Images in the Context of Visual Collectons and Scholarshp," *Visual Resources*, Vol X, 1990, pp. 11-24 and "Specifics of Imaging Practice," *Archives & Museum Informatics*, 1995, pp. 147-158.

[8] Roger S. Bagnall, *Digital Imaging of Papyri: A Report to the Commission on Preservation and Access*, Commission on Preservation and Access, September 1995; Janet Gertz, *Oversize*

Color Images Project, 1994-1995 Final Report of Phase 1, Commission on Preservation and Access, August 1995; Picture Elements, Inc., *Guidelines for Electronic Preservation of Visual materials, Part 1*, (2 March 1995), and Reilly. Michael Ester argues that an "archival image" of a photograph can not be benchmarked through calculations, but should be pegged to the "functional range of an institution's reproduction sources" see p. 11 in Ester, *Digital Image Collections: Issues and Practice*, Dec. 1996(CPA). For a critique of this approach, see Stephen Chapman and Anne R. Kenney, "Digital Conversion of Library Research Materials, A Case for Full Informational Capture," *D-Lib Magazine*, October 1996.

[9] Anne R. Kenney and Stephen Chapman, "Film Scanning," (Chapter Seven) in *Digital Imaging for Libraries and Archives*, June 1996, p. 169.

[10] ANSI/AIIM MS23-1991, *Practicefor Operational Procedures/Inspection and Quality Control of First-generaton, Silver Microfilm and Documents*, Association for Information and Image Management; ANSI/AIIM TR26-1993, *Resolution as it Relates to Photographic and Electronic Imaging*, Association for Information and Image Management; and Kenney and Chapman, *Tutorial: Digital Resolution Requirements for Replacing Text-Based Material: Methods for Benchmarking Image Quality*, Commission on Preservation and Access, April 1995.

[11] For a description of this verification process, see: Anne R. Kenney, "Digital-to-Microfilm Conversion: An Interim Preservation Solution," *Library Resources and Technical Services* (October 1993), pp. 380-401; (January 1994), pp. 87-95.

[12] A fuller explanation of the display benchmarking process is included in Kenney and Chapman, "Chapter 2", *Digital Imaging for Libraries and Archives* (June 1996), Cornell University Library, pp. 76-86.

[13] Improvements in managing color digitally may be forthcoming from an international consortium of industry leaders working to develop an electronic pre-press industry standard. Their "International Color Consortium Profile Format" is intended to represent color consistently across devices and platforms.

[14] See Peter van Minnen, "Imaging the Duke papyri," (December 1995) http://odyssey.lib.duke.edu/papyrus/texts/imaging.html, and Roger S. Bagnall, *Digital Imaging of Papyri: A Report to the Commission on Preservation and Access*, Commission on Preservation and Access, September 1995.

[15] Rhyne, *Computer Images for Research, Teaching, and Publication in Art History and Related Disciplines*, Commission on Preservation and Access, 1996, p. 5.

[16] The formula forcalculting the maximum percentage of a digital image that can be displayed on screen is:
a. If both image dimensions ≤ the correspondingpixel dimensions (pd) of the screen, 100% of the image will be displayed
b. If both image dimensions > the corresponding pixel dimensions of the screen,

$$\%\text{displayed} = \frac{\text{horiz. screen pd} \times \text{vertical screen pd}}{\text{Image's horiz. Pd} \times \text{image's vertical pd}} \times 100$$

c. If one of the image's dimensions ≤ the corresponding pixel dimension of the screen,

$$\%\text{displayed} = \frac{\text{screen'sopposite pixel dimension}}{\text{Image's opposite pixel dimension.}} \times 100$$

[17] The formula for scaling for complete display of image on screen is:
a. When digital image aspect ratio ≤ screen aspect ratio, set image's horizontal pixel dimension to the screen's horizontal pixel dimension
b. When digital image aspect ratio is > screen aspect ratio, set image's vertical pixel dimension to the screen's vertical pixel dimension.

[18] This formula presumes that bitonal images are presented with a minimum level of gray (3 bits or greater), and that filters and optimized scaling routines are used to improved image presentation.

……

For additional information about the conference, or The Andrew W. Mellon Foundation's scholarly communication initiatives, please contact Richard Ekman. For additional information about ARL or this web site contact Patricia Brennan, ARL Program Officer at (202) 296-2296.

Return to Office of Scholarly Communication Home Page

ARL Home ARL Scholarly Communication and Technology Home Page

© Association of Research Libraries, Washington, DC
Web Design by Angelo F. Cruz
Maintained by ARL Web Administrator
Last Modified: August 1, 1997

Scholarly Communication and Technology

Conference Organized by The Andrew W. Mellon Foundation

at Emory University
April 24-25, 1997

Copyright © of the papers on this site are held by the individual authors or The Andrew W. Mellon Foundation. Permission is granted to reproduce and distribute copies of these works for nonprofit educational or library purposes, provided that the author, source, and copyright notice are included on each copy. For commercial use, please contact **Richard Ekman** at the The Andrew W. Mellon Foundation.

Session #6 Copyright and Fair Use

Mellon Conference: Panel on "Licensing, Copyright and Fair Use" The HYPATIA Project (toward an ASCAP for Academics)

Jane C. Ginsburg and Morton L. Janklow
Professor of Literary and Artistic Property Law
Columbia University School of Law

* * *

[6/13/97]

The HYPATIA[1] Project (toward an ASCAP for Academics)

This project envisions the creation of a digital depository and licensing and tracking service for unpublished "academic" works, including working papers, other works-in-progress, lectures, and other writings that are not normally published in formal academic journals. A centralized digital clearinghouse for this material confers a number of benefits on the academic authors and on users, particularly users of university libraries, including students, professors, and other researchers.

First, a centralized depository offers a more systematic and convenient means to discover the unpublished literature than wandering around individual professors' or departments' Web pages. The depository's detailed and dynamic catalogue of its works, identifying new and revised submisions, will significantly enhance the accessibility of this material.

Second, academic authors may not always have a significant financial stake in the electronic exploitation of their works (whether the works are unpublished or published; in the latter instance, many academics may have assigned all rights to publishers -- sometimes inadvertently). But academics do have a very significant glory interest. A depository that undertakes what one might call "prestige accounting" for the authors, adds an important feature, and may serve as an incentive to participation.

What is "prestige accounting"? It is the tracking of use in a way that would permit authors to interrogate the depository to learn if and how their works are being used, for example on reserve or in coursepacks at identified universities, for identified courses. Currently, academic authors generally do not know, apart from general sales figures (if they receive any), what has been the dissemination of their works. With some prodding of publishers, one might find out which bookstores placed orders for the book, and thus infer which schools were using the work. However, this kind of information is not generally available (or, at any rate, disseminated) for photocopied course packs, even when rights are cleared.

Third, and especially important to the digital environment, a service of this kind would add considerable value if it could ensure that the digital version made available is authentic. Many works may be travelling on the Web, but the user may not (or should not) be confident that the document downloaded is completely consistent with the work as created. This is particularly significant when many different versions (e.g., prior drafts) are accessible at multiple Internet sites (not all of them with the author's permission).

I. Defining the *HYPATIA Universe*

A. What kinds of works will the HYPATIA depository include?

At least as an initial matter, the depository will be confined to unpublished works such as drafts, lectures, occasional pieces, conference proceedings, maseters theses, and, perhaps, doctoral dissertations. This definition should help avoid possible conflict with publishers (or those that are the copyright holders of works written by academics), who are or will be undertaking their own licensing programs. Moreover, the universe of "unpublished" works may grow as that of formal academic publications shrinks.

B. Whose works will be included in the HYPATIA depository?

Any academic [term to be defined; e.g., anyone with an institutional IP address] who wishes to deposit a work will be welcome to do so. There will be no screening or peer review.

Participating authors will register with the HYPATIA depository and will receive a password (registration information will also be relevant to terms and conditions, to authenticity; the password will tie into use reporting, see IIC; IVA; VB, *infra*).

II. Deposit

A. Entry of works

Deposits must be made by or under the authority of the author (if living) or her successor in title (if dead); the depository will not accept submissions from unauthorized third parties.

Deposited works should be sent in html format.[2]

Upon depositing, the author will supply information necessary to cataloguing the work, including her name and the title of the work, and will categorize the work for the HYPATIA catalogue by selecting from LC classifications and subclassifications supplied on menu screens (See also IIIC, *infra*.)

Every work deposited in HYPATIA will automatically receive an identifying ISBN-type number ("HYPATIA number"). The number will be communicated to each author upon deposit, as well as maintained in the catalogue.

B. Exit of Works

The author, upon submitting the work, may demand that it self-delete from the depository by a date selected. Any document so designated should bear a legend indicating when it will no longer be included in the depository.

The author may also demand deletion from the depository at any time. The catalogue (see IIIC, *infra*) will indicate when a work has been deleted as well as if it has been replaced by an updated version. A "morgue catalogue" will be established to keep a record of these deletions.

C. Terms and Conditions

With each deposit, a participating author who wishes to impose terms and conditions on use of the work may select from a menu of choices. These will include:

> What kind of access to permit (e.g., browsing only)
>
> What purpose (e.g., personal research but not library reserve or course packs)
>
> Whether or not to charge for:
> Access
> Storage
> Further reproductions
>
> [Additional terms and conditions to be provided]

III. Access

A. What Users May Access the HYPATIA Depository?

As a starting point, access will be limited to university-affiliated (or research institute-affiliated) users. These users will make their first contact with HYPATIA from their institutional host, in order to establish a user ID number from which they may subsequently gain access from both institutional and non institutional hosts (i.e. work or home).

When registering, the user will indicate her user category (e.g., Professor, post-doctoral, graduate, undergraduate) and disciplines (research and teaching subject matter areas); this information will be relevant to the depository's catalogue and tracking functions (see IIIC, VA, *infra*).

A second phase of the project would extend access to independent scholars who do not have institutional affiliations. At a later date, access to the depository might be expanded to the general public.

B. Conditions on Use

When registering, the user will encounter a series of screens setting forth the general conditions on using HYPATIA. These include agreement to abide by the terms and conditions (if any) each author has imposed on the deposited works. (E.g., the author permits browsing and personal copying, but not futher copying or distribution.) The user will also agree that in the event of a dispute between the user and HYPATIA, or between the user and a HYPATIA author, any judicial proceding will be before the U.S. District Court for the Southern District of New York (or, if that court lacks subject matter jurisdiction, before the New York State Supreme Court), and will be governed by U.S. copyright law and New York law.[3]

C. How Will Users Know What Are HYPATIA's holdings?

The depository will include an electronic catalogue searchable by key word or by Boolean logic. The catalogue will also be organized in a scroll-through format employing LC subject headings. The catalogue will be dynamic, so as to reflect new submissions, or revisions of material (and will also indicate when an author has deleted material from the depository).

The catalogue will be dynamic in another way. Along the lines of "SmartCLIP" and similar products, it will regularly e-mail registered users with information about new submissions in the subject matter categories the registrant has requested.

D. How would users access material from the HYPATIA depository?

After finding the requested work's HYPATIA number in the general online catalogue, or in the e-mailed updates, the registered user will click on the catalogue or type in the HYPATIA number to receive the work.

It is also possible to envision links to specific works in the depository from online course syllabi or other online reading lists.

In addition to the general conditions screens encountered on first registration, the terms and conditions (if any) pertinent to each work communicated will appear on the initial screen prefacing each work. In order to access the rest of the document, the user will be obliged to click on her consent to those terms and conditions.

IV. Authenticity

A. Delivery from the HYPATIA depository

Documents in the depository will be authentic when submitted by the author. The depository will add digital signatures or other marking material to identify the author, the work, and its date of submission.

B. Subsequent generations of documents originally obtained from the depository

The HYPATIA project does not now contemplate attempting to prevent users from making or circulating further copies of works obtained from the depository. But it is important to make it possible for anyone who obtains a document of uncertain provenance to compare it with the authentic version in order to ensure that no alterations have occurred. Thus, if a registered user has obtained a copy from a source other than HYPATIA, the user should verify that copy against the version in the depository.

V. Tracking

A. Identification of Uses

Registered users will respond to a menu screen indicating the purpose of their access: e.g., library reserve; coursepack; personal research.

B. Reporting

Registered authors will have electronic "prestige" reports that they may interrogate anytime to learn:

the number of "hits" each deposited work has received

the source of the hit (institution, department, user category -- names of users will not be divulged)

The nature of the use (library reserve; coursepack; research)

C. Billing

If the author has requested payment for access or copying, the registered user will need a debit account to access the work; the debit would be credited to the author's account. These operations may be implemented through links to a participating bank.

VI. Other Potential Applications of HYPATIA

As currently conceived, HYPATIA's universe is unpublished academic works. But once all its features have been put into place, HYPATIA could either expand its holdings, or work in tandem with copyright owners of published works to supplement whatever rights clearance system the publisher has devised. Similarly, where authors have not assigned their copyrights, or have at least retained electronic rights, HYPATIA could work together with collective licensing agencies, such as the Authors' Registry, to supplement their rights clearance and reporting mechanisms.

VII. Costs of Implementation and Maintenance

A. Initial Setup

The primary initial costs will be in acquiring hardware to accommodate the depository, and in creating or adapting the software for the various components of the system: author registration; deposit; cataloguing; user registration; use tracking and reporting; billing. It will also be important to publicize HYPATIA to potential participating institutions and authors and users; some portion of the initial budget should be allocated to this.

B. Maintenance

Because most of the information in HYPATIA is author- or user-generated, the maintenance costs should be largely limited to general system maintenance and gradual expansion of disk storage. It may be desirable to provide for part-time "help line" assistance.

C. Paying for HYPATIA

It will be necessary to seek a grant to support the initial setup of and publicity for the system. The maintenance and helpline costs should be covered by a modest subscription from participating institutions, in exchange for the service of receiving and delivering works into/from the depository.

If the payment feature becomes a significant aspect of HYPATIA, a portion of the access or copying charges could go to defray maintenance expenses.

This project was developed by Jane C. Ginsburg, Morton L. Janklow Professor of Literary and Artistic Property Law, Columbia University School of Law, in consulation with James Hoover, Professor of Law and Associate Dean for Library and Computer Services, Columbia University School of Law; Carol Mandel, Deputy University Librarian, Columbia University; David

Millman, Manager, Academic Information Systems, Columbia University; and with research assistance from Deirdre von Dorum, Columbia University School of Law, class of 1997.

The HYPATIA Project: Annotated Bibliography of Online Sources

---

FOOTNOTES:

[1] Hypatia was the patron of libraries; the librarians at Alexandria claimed descendence from her. Oxford Classical Dictionary 534 (1970). As an acronym, the name stands for "HTML Your Paper At This Internet Address." See IC, *infra*.

[2] Hence the acronym "HTML Your Paper At This Internet Address."

[3] The choice of forum and of state law assumes that HYPATIA will be established at Columbia University.

……

For additional information about the conference, or The Andrew W. Mellon Foundation's scholarly communication initiatives, please contact Richard Ekman. For additional information about ARL or this web site contact Patricia Brennan, ARL Program Officer at (202) 296-2296.

Return to Office of Scholarly Communication Home Page

ARL Home ARL Scholarly Communication and Technology Home Page

© Association of Research Libraries, Washington, DC

Web Design by Angelo F. Cruz
Maintained by ARL Web Administrator
Last Modified: August 29, 1997

The Transition to Electronic Content Licensing
The Institutional Context in 1997

Scholarly Communication and Technology Conference
of the Andrew. W. Mellon Foundation
Emory University
April 24-25, 1997

Ann Okerson
Yale University Library
Ann.Okerson@yale.edu

Introduction

The public discourse about electronic publishing, as heard at scholarly and library gatherings about the topic of scholarly communications, has changed little over the last several years. Librarians and academics fret about the serials crisis, argue about the influence of commercial off-shore publishers, wonder when the academic reward system will begin to take electronic publications into account, and debate what steps to take to rationalize copyright policy in our institutions. There is progress in that a wider community now comes together to ponder these familiar themes, but to those of us who have been party to the dialog for some years, the tedium of ritual sometimes sets in.

At Yale, subject-specialist librarians talk to real publishers every day about the terms on which the Library will acquire their electronic products: reference works, abstracts, data, journals, and other full-text offerings. Every week, or several times a week, we are swept up in negotiating the terms of licenses with producers whose works are needed by our students and faculty. In 1997, electronic publications are a vital part of libraries' business and services. For example, at a NorthEast Research Libraries Consortium (NERL) meeting in February, each of the 13 research library representatives at the table stated that its library is expending about 6-7% of its acquisitions budget on electronic resources.

This essay will offer some observations on the overall progress of library licensing negotiations. But the main point of the will be to make this case: in the real world of libraries, we have begun to move past the predictable, ritual discourse. The market has brought librarians and publishers together; the parties are discovering where their interests mesh; and they are beginning to build a new set of arrangements that meet needs both for access (on the part of the institution) and remuneration (on the part of the producer). Even though the prices for electronic resources are becoming a major concern, libraries are able to secure crucial and significant use terms via site licenses, terms that often allow the customer's students, faculty, and scholars significant copying latitude for their work (including articles for reserves and coursepacks), at times more than what permitted via the fair use and library provisions in the Copyright Act of the U.S. In short, institutions and publishers are as or more advanced in making a digital market than perhaps they realize and more advanced than they are with resolving a number of critical technological issues.[1]

Why do Contracts or Licenses (Rather Than Copyright) Govern Electronic Content?

Society now faces what seems to be a powerful competitor for copyright's influence over the marketplace of cultural products, one that carries its own assumptions about what intellectual property is, how it is to be used, how it can be controlled, and what economic order can emerge as a result.

For convenience's sake, the codification of intellectual property is assigned to the early eighteenth century.

That is when the evolving notion of copyright was enacted into law, shaping a marketplace for cultural products unlike any seen before. In that 18th century form, copyright legislation depended in three ways on the technologies of the time.

- First, the power of copyright was already being affirmed through the development of high-speed printing presses that increased the printer's at-risk capital investment and greatly multiplied the number of copies of a given original that could be produced (and thus lowered the selling price).
- Thus an author could begin to realize financial rewards through signing over copyright to a publisher. Owning the copyright meant that the publisher who had assumed the expense and risk of publication stood to gain a substantial portion of the publication revenue.
- Third, punishment for breaking the law (i.e., printing illegal copies) was feasible, for the ability to escape detection was relatively slight. The visibility and the capital costs of establishing and operating a printing press meant that those who used such presses to violate copyright were liable to confiscatory punishment at least commensurate with the injury done by the crime itself.

In the 1970s, technology advances produced the photocopier, an invention that empowered the user to produce multiple copies cheaply and comparatively unnoticed. In the 80s, the fax machine took the world by storm, multiplying copies and speeding up their distribution. Computer networking technology of the 90s marries the convenience, affordability, and ease of distribution, eclipsing the power of all previous technologies. We can attribute the exponential increase in electronic content, at least indirectly, to the current inhabitants of the White House. The Clinton-Gore campaign of 1992 placed the Internet before the general public and this administration has been passionately committed to rapid development of the National Information Infrastructure (NII) and determined to advance the electronic marketplace. Part of that commitment arises from national leader's unwavering faith that electronic networks create an environment and a set of instruments vital to the overall economic growth of the United States.

While copyright (that is, the notion that creative works can be owned) is still and probably always will be recognized as a fundamental principle by most players in the information chain, many believe that its currently articulated "rules" do not effectively address either the technical capabilities or reader needs of a high-speed information distribution age. And while it could be argued (and many educators do) that the 19th and 20th century drafters of copyright law intended to lay down societally-beneficial, and by extension technologically-neutral, principles about intellectual property ownership and copying,[2] in fact Thomas Jefferson knew nothing of photocopiers and the legislators who crafted the 1976 Copyright Act of the United States knew nothing of computer networks. There is a case to be made that, had they even begun to imagine such things, the law might have been written differently -- and that in fact it should now be written differently.[3] So, the gulf between copyright laws or treaties and the universe that those laws ought to address today, feels to many vast and deep. Therefore, instead of relying on national copyright law, surrounding case law, international treaties, and prevailing practice to govern information transactions for electronic information, copyright holders have turned to contracts (or licenses as they are more commonly called in the library world) as the mechanism for defining the owner, user, and uses of any given piece of information.

That is, the license-contract is invoked because the prospective deal is a substantial (in cash or consequence) transaction for both parties, feels like new kind of marketplace (or a market for a new kind of product) , and neither the selling or buying party is sure either of the other or of their position vis a vis the law or the courts. Publishers come to the table with real anxieties that their products may be abused by promiscuous reproduction of a sort that ultimately saps their product's marketability, while libraries are fearful that restrictions on permitted uses will mean less usable or more expensive products.

In short, what licensing agreements have in common with the copyright regime is that both accept the fundamental idea of the nature of intellectual property -- that even when intangible, it can be owned. Where they differ is in the vehicle by which they seek to balance creators, producers, and users rights and to regulate the economy that springs up around them. Copyright represents a set of general regulations negotiated through statutory enactment. Licenses on the other hand represent a market-driven approach to this regulation, through deals struck between buyers and sellers.

When Did This Mode of Doing Business Begin for Libraries?

The concept of a license is old and fundamentally transparent. A license is essentially a means of providing use of a piece of property without giving up the ownership. For example, if one owns a piece of property and allows another to use it without transferring title, one may by law of contract stipulate the conditions one chooses; if the other party agrees to them, then a mutually agreeable deal has come into being. A similar transaction takes place in the case of performance rights for films and recordings. In such an example, we move from the tangible property mode of real estate where exclusive licenses (granting of rights to only one user) are common, to the intangible property mode of intellectual property such as copyright -- where non-exclusive licenses are the norm. The owner of a movie theater rarely owns the cans of film delivered weekly to the cinema, holding them instead under strict conditions of use: so many showings, so much payment for each ticket sold, etc. As with the economic relationship between author and publisher that is sanctioned by copyright, with the right price such an arrangement can be extraordinarily fruitful. In the license mode of doing business (precisely defined by the legal contract that describes the license) the relationships are driven entirely by contract law: the owner of a piece of property is free to ask whatever price and set whatever conditions on use the market will bear. The ensuing deal is pure "marketplace": a meeting of minds between a willing buyer and a willing seller.

A crucial point here is that where the owner of the property has a copyright-protected monopoly, the license becomes a particularly powerful tool for that owner.

Most of academics began to be parties to license agreements when personal computer software (*WordStar, WordPerfect*) appeared in the 1980s in shrinkwrap packages for the first time. Purchasers of such software may have read the fine print on the wrapper detailing the terms and conditions of use, but for the most part they either did not or have ceased to do so. The thrust of such documents is simple: by opening the package, the purchaser has agreed to certain terms, terms that include limited rights of ownership and use of the item paid for. In many ways, this mode of licensing raises problematic questions,[4] but in others such as sheer efficiency, it suggests the kind of transaction that the scholarly information marketplace needs to achieve. It is noteworthy that the shrinkwrap license has moved easily into the World Wide Web environment, where it shows itself in clickable "I agree" form. The user's click supposedly affirms that he or she has said yes to the user terms and is ready to abide by them. The downsides and benefits are similar to those of shrinkwrapped software.

The phenomenon of institutional licensing for electronic content has evolved in a short time. Over the last 20 years or so, licensing software has become a way of life for institutions of higher education. These kinds of licenses are generally for systems such as those that run institutional computers or online catalogs or software packages (e.g., for instruction, office support). The licenses, often substantial in scale and price, are arranged by institutional counsel (an increasingly overworked segment of an educational institution's professional staff) along with information technology managers.

Libraries' entree into this arena has been comparatively recent and initially on a small scale. In fact, the initial library business encounter with electronic content may not have happened via license at all, but rather via deposit account. Some 20 years ago, academic and research libraries began accessing electronic information, at that time primarily through mediated searching of indexing and abstracting services through consolidators such as Dialog. Different database owners levied different per-hour charges (each database also required its own searching vocabularies and strategies), and Dialog (in this example) aggregated them for the educational customer. For the most part, libraries established accounts to which these (usually mediated by librarians or information specialists) searches were charged.

By the late 80s, libraries also began to purchase shrinkwrapped ("pre-licensed") content, though shrinkwrapped purchases did not form -- and still do not -- any very visible part of library transactions. Concurrently, a number of indexing and abstracting services offered electronic versions directly to libraries via CD-ROM or through dial-up (for example, an important early player in this arena was ISI, the Institute for Scientific Information), and it was at this point, within the last ten years, that library licenses gradually became recognized as a means to a new and different sort of information acquisition or access. Such licenses were often arranged by library subject specialists for important resources in well-defined areas of use. The license terms offered to libraries were accepted or not, the library customer regarding them mostly as non-negotiable. Non-acceptance was more often than not a matter of affordability, and there seemed to be little room for the library customer to affect the terms. Complaints about terms of licenses began to be (and persist in being) legion, for important reasons such as:

- *Potential loss of knowledge.* By definition, licenses are arranged for specific periods of time. At the end of that time, librarians rapidly discovered, if the license is not renewed, prior investment can become worthless as the access ceases (for example, a CD-ROM must be returned or perhaps it stops being able to read the information; connections to a remote server are severed).
- *License restrictions on use and users.* Institutions are often asked to assure that only members of the institution can use electronic information, in order to reduce or curtail its leakage.
- *Limitations on users' rights.* Initial license language not infrequently asks that institutional users severely limit what and how much they may copy from the information resource and may prescribe the means by which such copying can be done.
- *Cost.* In general, electronic licenses for indexing and abstracting services appeared, and still appear, to cost significantly more than print equivalents.[5]

What Has Happened to Increase Libraries' Awareness of Licenses?

1. *Sheer numbers.* Whatever their marketplace insecurities may be, thousands of information providers have jumped into the scholarly marketplace with electronic products of one sort or another: CDs, online databases, full text resources, multi-media. Many learned societies, scientific publishers, university presses, full-text publishers, vendor/aggregators, as well as new entrants to the publishing arena, now offer either beta or well-tested versions of either print-originating or completely electronic information. The numbers have ballooned in a short 2 to 3 years with no signs of abating. For example, *NewJour*, the online forum for announcing new e-journals, magazines and newsletters reports 3634 titles in its archive as of April 5, 1997, this without the 1100 science journal titles that Elsevier is now making available in electronic form.[6] The Yale University Library licenses over 400 electronic resources of varying sizes, types, media, and price and reviews about two new electronic content licenses a week.

2. *The attempt of various players in the information chain to create guidelines about electronic fair use, for example in the CONFU process[7] have not so far proved fruitful.* In connection with the Clinton Administration's National Information Infrastructure initiative, the Working Group on Intellectual Property Rights in the Electronic Environment called upon copyright stakeholders to negotiate guidelines for the fair use of electronic materials in a variety of nonprofit educational contexts. Anyone who wished to participate was invited to do so and a large group calling itself CONFU, the Conference on Fair Use, began to negotiate such guidelines for a variety of activities (such as library reserves, multimedia in the classroom, interlibrary loan, etc.) in September 1994. The interests of all participants in the information chain were represented, and

the group quickly began to come unstuck in reaching agreements on most of the dozen or more areas defined as needing guidelines. Such stalemates should come as no surprise; in fact, they are healthy and proper. Any changes to national guidelines, let alone national law or international treaty, should happen only when the public debate has been extensive and consensus can be reached. What many have come to realize during the current licensing activities is that the license arrangements that libraries currently are making, are in fact achieving legislation's business more quickly and by other means. Instead of waiting on Congress or CONFU and allowing terms to be dictated to both parties by law, publishers and institutions are starting to make their peace together, thoughtfully and responsibly, one step at a time. Crafting these agreements and relationships is altogether the most important achievement of the licensing environment.

3. *Numerous formal partnerships and informal dialogs have been spawned by capabilities of new publications technologies.* A number of libraries collaborate with the publishing and vendor communities as product developers or testers. Such relationships are fruitful in multiple ways. With regard to licensing, they encourage friction, pushback, and conversation that leads to positive and productive outcomes. Close to home, libraries have been offered -- and have greatly appreciated -- the opportunity to discuss at length the library licenses of various producers at this conference, JSTOR specifically, and libraries feel they have had the opportunity to shape and influence these with mutually satisfactory results.

4. *Library consortia have aggressively entered the content negotiating arena.* While library consortia have existed for decades, and one of their primary aims has been effective information sharing, it is only in the 90s (and mostly in the last 2 to 3 years) that a combination of additional state funding (for state-wide consortia), library demands, and producers' willingness to negotiate with multiple institutions have come together to make the consortial license an efficient and perhaps cost-effective way to manage access to large bodies of electronic content. An example of a particularly fruitful marketplace encounter (with beautiful as well as charged moments) occurred from February 3-5, 1997, as a group of consortial leaders, directors, and coordinators who communicate informally for a year or two through listserve messages, arranged a meeting at the University of Missouri-St. Louis. The Consortium of Consortia (COC, as we sweepingly named ourselves) invited a dozen major electronic content vendors to describe their products briefly and their consortial working arrangements in detail.[8] By every account, this encounter achieved an exceptional level of information-swapping, interaction, and understandings both of specific resources and of the needs of producers and customers. That said, the future of consortial licensing is no more certain than for individual library licenses, though for different reasons.[9]

5. *Academia's best legal talent offer invaluable support to libraries.* Libraries are indebted to the intelligent and outspoken lawyerly voices in institutions of higher learning in this country. The copyright specialists in universities' general counsel offices have in a number of cases led in negotiating content licenses for the institution and have shared their strategies and knowledge generously. Law school experts have published important articles, taught courses, contributed to internet postings, and participated in national task forces where such matters are discussed. [10]

6. *The library community has organized itself* to understand the licensing environment for its constituents. The Association of Research Libraries (ARL) has produced an introductory licensing brochure,[11] The Council on Library Resources/Commission on Preservation and Access has supported Yale Library's creation of an important WWW site about library content licensing,[12] and the Yale Library offers the library, publisher, vendor, and lawyer world a high-quality moderated online list where the issues of libraries and producers are aired daily.[13]

7. *There is no other way.* Licensing and contracts are the only way to do business right now for an increasing number of electronic information resources that Library users need for their education and research.

Some Notable Challenges of the Library Licensing Environment Today

I identify these because they are important and need to be addressed, treating this conference as a place to pose the questions in order that we may begin answering them.

1. *Terms of use.* This area needs to be mentioned at the outset, as it has caused some of the most anguished discussions between publishers and libraries. Initially, many publishers' contact language for electronic information was highly restrictive about both Permitted Users and Permitted Uses. Assumptions and requirements about how use ought to be contained have been at times ludicrous, for example, in phrases such as "no copies may be made by any means electronic or mechanical." Through dialog between librarians and producers, who are usually genuinely eager to market their work to happy customers, much of this language has disappeared from the first draft contracts presented to library customers. Where libraries are energetic and aggressive on behalf of their users, the terms of use can indeed be changed to facilitate educational and research goals. The Yale Library, for example, is now party to a number of licenses that permit substantial amounts of copying and downloading for individual learning, research, in-the-classroom learning, library reserves, coursepacks, and related activities. Interlibrary Loan and transmission of works to individual scholars in other organizations are matters that still need a great deal of work. However, the licenses of 1996 and 1997 represent significant all-around improvements and surely reinforce the feeling that rapid progress is being made.

2. *Scalability.* Institutional electronic content licenses are now generally regarded as negotiable, mostly because the library-customer side of the marketplace is now treating them as such (which publishers seem to welcome) and successes of different sorts have ensued (success being defined as a mutually agreeable contract), making all parties feel that they can work together effectively in this new mode. However, negotiations are labor-intensive. Negotiation requires time (to develop the expertise and to negotiate), and time is a major cost here. The current mode of one on one negotiations between libraries and their publishers seems at the moment necessary, for many reasons, and at the same time it places new demands on institutional staff. Scalability is the biggest challenge for the licensing environment.

- Clearly, it is too early to shift the burden onto intermediaries such as subscription agencies or other vendors who have vested interests of their own. So far their intervention has been absent or not particularly successful. In fact, in some of the situations where intermediaries purvey electronic databases, library customers secure less advantageous use terms than those libraries could obtain by licensing directly from the publishers. This is hardly surprising, as those vendors are securing commercial licenses from the producers, while libraries are able to obtain educational licenses. Thus, it is no surprise that in unveiling their latest electronic products and services, important organizations such as Blackwell's ("Navigator") and OCLC ("EJO - Electronic Journals On-line") leave license negotiating for the journal content as a matter between the individual journal publishers and their library customers.
- The contract that codifies the license terms is a pervasive document covering every aspect of the library/producer relationship, from authorized uses and users to technology base, duration, security mechanisms, price, liability, responsibility, etc. That is, the license describes the full dimensions of the "deal" for any resource. Attempts on the part of the library and educational community to draft general principles or models to address content licensing characteristically forget this important fact and the results inevitably fall short in the scaling-up efforts.

3. *Price.* Pricing models for electronic information are in their infancy; they tend to be creative, complicated and often hard to understand. [14] Some of these can models range from wacky to bizarre. Consortial pricing can be particularly complex. Each new model solves some of the equity or revenue problems associated with

earlier models but introduces confusion of its own. While pricing of electronic resources is not strictly speaking a problem with the license itself, price has been a major obstacle in making electronic agreements. The seemingly high price tags for certain electronic resources leave the "serials crisis" in the dust.[15] It is clear that academic libraries, particularly through their consortial negotiators, expect bulk pricing arrangements, sliding scales, early signing bonuses, and other financial inducements that publishers may not necessarily feel they are able to offer. Some of the most fraught moments at the St. Louis COC meeting involved clashes between consortial representatives who affirmed that products should be priced at whatever a willing buyer can or will pay, even if this means widely inconsistent pricing by the vendor, and producers who affirmed the need to stick with a set price that enables them to meet their business plan.

4. *The Liability-Trust Conundrum.* One of the most vexing issues for producers and their licensees has been the producer's assumption that institutions can and ought to vouch for the behavior of individual users (in licenses the sections that deal with this matter are usually called Authorized or Permitted Users and what Users may do under the terms of a license is called an Authorized or Permitted Use) and that individual users' abuses of the terms of a license can, in fact, kill the deal for a library or a whole group of libraries. Working through this matter with provider after provider in a partnership/cooperative approach poses many challenges. In fact, this matter may be a microcosm of a larger issue: the development of the kind of trust that must underlie any electronic content license. Generally the marketplace for goods is not thought of in terms of trust; it regarded as a cold-cash (or virtual cash) transaction environment. Yet the kinds of scaled-up scholarly information licenses that libraries are engaging with now depend on mutual understanding and trust in a way not needed for the standard trade -- or even the print -- market to work. In negotiating electronic content licenses, publishers must trust -- and, given the opening up of user/use language, it seems they are coming to trust -- their library customers to live up to the terms of the deal.

In part, we currently rely on licenses because publishers do not trust users to respect their property and because libraries are fretful that publishers will seek to use the new media to tilt the economic balance in their favor. Both fears are probably overplayed. If libraries continue to find, as they are beginning to do, that publishers are willing to give *the same or even more* copying rights via licenses as copyright offers, both parties may not be far from discovering that fears have abated, trust has grown, and the ability to go revert to copyright the primary assurance of trust can therefore increase. But many further technological winds must blow -- for example the cybercash facility to allow micropayment transactions -- before the players may be ready to settle down to such a new equilibrium.

5. *The Aggregator Aggravation (and Opportunity).* The costly technological investments that producers need to make to move their publications onto an electronic base; the publishing processes that are being massively re-conceived and reorganized; and not least, the compelling vision of digital libraries that proffer information to the end user through a single or small number of interfaces, with a single or modest number of search engines, gives rise to information aggregators of many sorts: [16] those who develop important searching, indexing, and/or display softwares (AltaVista, OpenText, etc.); those who provide an interface or gateway to products (Blackwell, etc.), and those who do all that plus offer to deliver the information (DIALOG@CARL, OCLC, etc.). Few publishers convert or create just one journal or publication in an electronic format. From the viewpoint of academic research libraries, it appears that the electronic environment has the effect of shifting transaction emphasis from single titles to collections or aggregations of electronic materials as marketplace products.

In turn, licensing collections from aggregators makes libraries dependent on publishers and vendors for services in a brand new way. That is, libraries' original expectation for electronic publications, no more than five years ago, was that publishers would provide the data and the subscribing library or groups of libraries would mount and make content available. But mounting and integrating electronic information requires a great deal of capital, effort, and technological sophistication, as well as multiple licenses for software and content. Thus, the prognosis for institutions meeting all or most of their users' electronic information needs locally is slim. The currently emerging mode, thus, takes us to a very different world in which publishers have positioned themselves to be the electronic information providers of the moment.[17]

The electronic collections offered to the academic library marketplace are frequently not in configurations librarians would have chosen for their institutions, had these resources been unbundled. This has been an issue in several of Yale Library's negotiations. Say that the publisher of a large number of quality journals makes only the full collection available in e-form, and only through consortial sale. By this means, the Yale Library recently "added" 50 electronic journal titles to its cohort, titles it had not chosen to purchase in print. The pricing model did not include a cost for those additional 50 titles; it was simply easier for the publisher to include all titles than to exclude the less desirable ones. While this paper is not the place to explore this particular kind scaling up of commercial digital collections, I leave it as a topic of potentially great impact on the academic library world.

6. *The Challenge of Consortial Dealings.* Ideally, groups of libraries acting in consort to license electronic resources can negotiate powerfully for usage terms and prices with producers. In practice, both licensors and licensees have much to learn about how to this scaled up environment. Some of the particularly vexing issues, for example, include:

- Not all producers are willing to negotiate with all consortia; some are not able to negotiate with consortia at all.
- In the early days of making a consortial agreement, the libraries may not achieve any efficiencies because all of them (and their institutional counsel) may feel the need or desire to participate in the negotiating process. Thus, in fact, a license for 12 institutions may take nearly as long to negotiate as 12 separate licenses.
- Consortia overlap greatly, particularly with existing bodies such as cataloging and lending "utilities" offering consortial deals to their members. It seems that every library is in several consortia these days, and many of us are experiencing a "competition" for our business from several different consortia at once for a single product's license.
- No one is sure precisely what a consortial "good deal" comprises. That is, it is hard to define and measure success. The bases for comparison between individual institutional and multiple institutional prices are thin and the stated savings can often feel like a sales pitch.
- Small institutions are more likely to be unaffiliated with large or powerful institutions and left out of seemingly "good deals" secured by the larger, more prosperous libraries. Surprisingly enough, private schools can be at a disadvantage since they are generally not part of state-established and funded consortial groups.
- In fact, treating individual libraries differently to collectives may, in the long run, not be in the interests of publishers or those libraries.

7. *Institutional Workflow Restructuring.* How to absorb the additional licensing work (and create the necessary expertise) within educational institutions is a challenge. One can foresee a time when certain kinds of institutional licenses (electronic journals, for example) might offer fairly standard, signable language, for surely producers are in the same scaling-up bind that libraries are. At the moment, licenses are negotiated in various departments and offices of universities and libraries. Many universities require that license negotiation, or at least a review and signature, happen through the Office of General Counsel, and sometimes over the signature of the Purchasing Department. In such circumstances, the best result is delay; the worst is that the Library may not secure the terms it deems most important. Other institutions delegate the negotiating and signing to library officers who have an appropriate level of responsibility and accountability for this type of legal contract. Most likely the initial contact between the Library and the electronic provider occurs by the public service or collections librarians who are most interested in bringing

the resource to campus.

One way of sharing the workload is to make sure that all selector staff receive formal or informal training in the basics and purposes of electronic licenses, so that they can see the negotiations through as far as possible, leaving only the final review and approval to those with signing authority. [18] In some libraries, the licensing effort is coordinated from the Acquisitions or Serials Departments, the rationale being that this is where purchase orders are cut and funds released for payment. However, such an arrangement can have the effect of removing the publisher interaction from library staff best positioned to understand a given resource and the library readers who will be using it. Whatever the delegation of duties may be at any given institution, it is clear that the tasks must be carved out somewhere in a sensible fashion, for it will be a long time before the act of licensing electronic content becomes transparent. Clearly, this new means of working is not the "old" acquisitions model. How does everyone in an institution who should be involved in crafting licensing "deals" get a share of the action?

Succeeding (Not Just Coping)

On the positive side, both individual libraries and consortia of libraries have reported negotiating electronic content licenses with a number of publishers who have been particularly understanding of research library needs. In general, academic publishers are proving to be willing to give and take on license language and terms, provided that the licensees know what terms are important to them. In many cases, librarians ask that the publisher re-instate the "public good" clauses of the Copyright Act into the electronic content license, allowing fair use copying or downloading, interlibrary loan, and archiving for the institutional licensee and its customers. Consortial negotiations are having a highly positive impact on the usefulness and quality of licenses.

While several downsides to the rapidly growing licensing environment have been mentioned, the greatest difficulty at this point is caused by the proliferation of licenses that land on the desks of librarians, university counsel, and purchasing officers. The answers to this workload conundrum might lie in several directions.

1. *National or Association Support.* National organizations such as ARL and The Council on Library Resources are doing a great deal to educate as many as possible about licensing. Practising librarians treasure that support and ask that licensing continue to be part of strategic and funding plans. For example, the Yale Library has proposed next-step ideas for the World Wide Web Liblicense project and appreciate the Council's interest in them. Under discussion are such possibilities as: further development of a prototype licensing software that will enable librarians to create licenses on the fly, via the World Wide Web, for presentation to producers and vendors as a negotiating position[19] and assembling a working group meeting that involves publisher representatives in order to explore how many pieces of an academic electronic content are amenable to standardization. Clearly, academic libraries are working with the same producers to license the same core of products over and over again. It might be valuable for the ARL and other organizations to hire a negotiator to develop acceptable language for certain key producers, say the top 100, with the result that individual libraries would not need to work out this language numerous times. Pricing and technology issues, among others, might nonetheless need to remain as items for local negotiation.

2. *Aggregators.* As indicated above, as libraries, vendors, and producers become more skilled as aggregators, the scaling issues will abate somewhat. Three "aggregating" directions are emerging:

- Information Bundlers, such as Lexis-Nexis, OCLC, DIALOG@CARL, UMI, IAC, OVID, and a number of others offer large collections of materials to libraries under license. Some of these are sizeable take-it-or-leave-it groupings; others allow libraries to choose subsets or groups of titles.
- Subscription Agents are beginning to develop gateways to electronic resources and to offer to manage

libraries licensing needs.

- Consortial of Libraries can be considered as "aggregators" of library customers for publishers.

3. *Transactional Licensing*. This paper treats only institutional licenses, be they site licenses, simultaneous user/port licenses, or single-user types. An increasing number of library transactions demand rights clearance for a piece at a time (situations that involve, say, course reserves or provision of articles that are not held in the library through a document supplier such as CARL). Mechanisms for easy or automatic rights clearance are of surpassing importance and various entities are applying considerable energies to them. The academic library community has been skittish about embracing the services of rights management or licensing organizations, arguing that participation would abrogate fair use rights. It seems important, particularly in light of recent court decisions, that libraries pay close attention to their position vis a vis individual copies (when they are covered by fair use and when they are not, particularly in the electronic environment) and take the lead in crafting appropriate and fair arrangements to simplify the payment of fees in circumstances when such fees are necessary.[20]

Beyond the License?

As we have seen, the content license comes into play when the producer of an electronic resource seeks to define a "deal" and an income stream to support the creation and distribution of the content. Yet, other kinds of arrangements are possible.

1. *Unrestricted and For Free*. One hears in many venues of important resources funded up front by governments or institutions, say, and the resources are available to all end users. Some examples include the notable Los Alamos High Energy Physics Preprints; the various large genome databases; the recent announcement by the National Institutes of Health of MEDLINE availability online; and numerous university-based electronic scholarly journals or databases. A number of such important resources exist and their numbers are growing, though they may always be in the minority of scholarly resources. Characteristically, such information is widely accessible, the restrictions on use are minimal or non-existent, and license negotiations are largely irrelevant or very straightforward.

2. *For a Subscription Fee and Unrestricted to Subscribers*. Some producers are, in fact, charging an online subscription fee but licenses need not be crafted or signed. The terms of use are clearly stated and generous. The most significant and prominent example of such not-licensed but paid-for resources include the rapidly growing collection of high-impact scientific and medical society journals published by Stanford University's HighWire Press.[21]

Both of these trends are important; they bear watching and deserve to be nurtured. In the first case, the up front funding model seems to very well serve the needs of large scientific or academic communities without directly charging users or institutions; they are products of public- or university-funded research. In the second instance, although users are paying for access to the databases, the gap between the copyright and licensed way of doing business seems to have narrowed and in fact the HighWire publications are treated as if copyright-governed. Over time, it would not be unreasonable to expect this kind of merger of the two (copyright and contract) constructs and to benefit from the subsequent simplification the merger would bring.

In short, there is much still to be learned in the content licensing environment, but much has been learned already. We are in a period of experimentation and exploration. All the players have real fears about the security of their livelihood and mission; all are vulnerable to the risks of information in new technologies; many are learning to work together pragmatically towards at least mid-term modest solutions, in turn using those modest solutions as stepping stones into the future.

NOTES

1. Clifford Lynch in "Technology and its Implications for Serials Acquisitions," *Against the Grain* 9:1 (1997), pp. 31+. This is a version of a talk by Lynch at the November 1996 Charleston Conference. He identifies they key needs in building digital libraries as authentication, printing, individual item addressability, accessibility, and linkage. Lynch concludes with this insight: "The theme I want to underscore here is that we need to be very careful about whether we have technology that can deliver this electronic content *for which we are busy negotiating financial arrangements in acceptable ways on a broad systemic basis.*" [emphasis is mine]

2. The statement "Fair Use in the Electronic Age: Serving the Public Interest," is an outgrowth of discussions among a number of library associations regarding intellectual property, and in particular, the concern that the interests and rights of copyright owners and users remain balanced in the digital environment. This important position statement was developed by representatives of the following associations: American Association of Law Libraries, American Library Association, Association of Academic Health Sciences Library Directors, Association of Research Libraries, Medical Library Association Special Libraries Association. It espouses the philosophy that the US copyright law was created to advance societal goals and well-being and embeds the notion of technological neutrality. It can be found at: gopher://arl.cni.org:70/00/scomm/copyright/policy/uses.

3. Close to home, I have recently had the opportunity to read statements from the international publishing community in two major position papers originating with the International Publishers Copyright Council, the STM group of publishers, and the International Publishers Association.
These documents affirm the following kinds of things:

- Digital versions of works are not the same as print versions, because digital information can be manipulated and widely distributed. (The implication is that all of this will happen and that it is happening with copyrighted works, often in an illegal manner.)
- Digital versions of works need even more protection than printed versions.
- Digital browsing is not the same as reading print: the very act of browsing involves reproducing copies (which immediately implicates and possibly violates copyright law).
- There should be no private or personal exemptions from copyright in the digital environment.
- There should be no exceptional copyright treatment for libraries in the digital environment -- the exemptions for traditional materials, if carried over into the digital environment, will result in unfair competition with publishers.
- Digital lending (a digital analog to ILL) will destroy publishers.
- Publishers are now poised to offer and charge for electronic delivery of information and therefore they ought to be able to do this. Such services will replace most of the copying libraries and individuals used to do in print.
- The role of libraries will be to provide access, select materials for users via what they choose to license, instruct users in the vast array of electronic sources and how to use them; support them in searching and research and learning needs.

4. See the recent decision *ProCD v. Zeidenberg* In the United States Court of Appeals For the Seventh Circuit, June 30, 1996. The question posed was: Must buyers of computer software obey the terms of shrinkwrap licenses? The district court had said not. The 7th Circuit reversed this decision. ProCD (plaintiff) compiled information from 3,000+ phone directories into one database, with additional information such as zip code extensions) and with their own searching software. They packaged it as a CD for personal sale in a shrink-wrap box. They also sold it in other ways to commercial companies as mailing lists and so on. On the

basis that factual information cannot be copyrighted, Mr. Zeidenberg bought a package of SelectPhone (TM) at a shop in Madison WI. He formed a company to re-sell the information which he made available over the WWW apparently quite cheaply. Zeidenberg argued that one cannot be bound by the shrinkwrap license because the terms are not known at the time of purchase. They are inside the package and the purchaser cannot be bound by terms that are secret at time of purchase. The Judges' Decision was that the shrinkwrap license is legal and a buyer is bound by it. The full decision and rationale may be read at the following URL: http://www.sgpdlaw.com/cases/procd_op.html

5. See Martha Kellogg, "CD-ROM Prod s as Serials: Cost Considerations for Libraries," in *Serials Review* 17:3 (1991), pp. 49-60. Using the tables in this article as a basis of comparison between print reference or I&A works and their CD equivalents shows a difference of about 30% where resources are comparable. Recent e-mail from the University of Michigan Library suggests differentials between print and electronic as high as 60%.

6. *NewJour* is a joint project of the Yale Library, the University of Pennsylvania, and the UC San Diego Library. Its fully searchable archive is located at: http://gort.ucsd.edu/newjour/

7. A good summary of the flavor, debates, and progress of CONFU can be found at URL: http://www.utsystem.edu/OGC/IntellectualProperty/confu.htm The CONFU interim report is available at URL: http://www.uspto.gov/web/offices/dcom/olia/confu/

8. For a list of the consortia that participated in the St. Louis meeting and descriptions of their activities, see the COC home page maintained by Bonnie Turner, Senior Administrative Assistant at Yale University Library: http://www.library.yale.edu/ocshelve/

9. Ann Okerson, "Buy or Lease? Two Models for Scholarly Information at the End (or the Beginning) of an Era), *Daedalus* 125:4 (1996), pp. 55-76. This is the special issue on libraries called Books, Bricks, and Bytes. I suggest that one possible outcome of the new trend to scaled-up consortial licensing activities is that the library marketplace will gain significant power and that publishers of scholarly information could find themselves in quite a different position than the "captive" marketplace of today. It is possible to argue that such an outcome is very healthy; on the other hand, even librarians and scholars might find it undesirable in that it would put today's specialized scholarly publications, with their attendant high prices, out of business. It seems to me that such publications are already most at risk as commercial (i.e., "for sale") publications and that they offer a perfect opportunity for scholars, universities, and libraries to devise a different mode of publication and distribution. The *Daedalus* piece can also be found at the URL: http://www.library.yale.edu/~okerson/daedalus.html

10. For example, the University of Texas Office of General Counsel's Copyright Management Center's site is an especially rich resource. The Center provides guidance and information to faculty, staff and students concerning applicable law and the alternatives available to help accomplish educational objectives. A large number of materials are accessible through the Web site, organized by topic. Some important documents stored directly on the Web server. The principal author is Georgia Harper, Copyright Counsel for the University of Texas System. The URL is: http://www.utsystem.edu/ogc/intellectualproperty/cprtindx.htm. Among others, the higher education community is indebted to Indiana University's Kenneth Crews, an important voice in CONFU (see for example the CETUS Fair Use document at: http://www.cetus.org/fairindex.html); the University of North Carolina Law School's Lolly Gasaway, also a leader in CONFU and contributor of many important resources (see for example "When Works Pass Into the Public Domain" at URL: http://www.library.yale.edu/~okerson/pubdomain.html); and Karen Hersey of the MIT Counsel's Office, a leader in crafting university-producer electronic license agreements and a frequent workshop presenter on this topic.

11. See "Licensing Electronic Resources: Strategic and Practical Considerations for Signing Electronic Information Delivery Agreements" at URL: http://arl.cni.org/scomm/licensing/licbooklet.html

12. See LIBLICENSE; Licensing Digital Information -- A Resource for Librarians. This Web resource contains license vocabulary, licensing terms and descriptions, sample publishers licenses, links to other licensing sites, and a bibliography about the subject. The URL is: http://www.library.yale.edu/~llicense/index.shtml

13. LIBLICENSE-L is a moderated list for the discussion of issues related to the licensing of digital information by academic and research libraries. To join the Liblicense-l list, please do the following: Send a message to: listproc@pantheon.yale.edu Leave the subject line blank. In the body of the message, type: subscribe LIBLICENSE-L Firstname Lastname

14. A LIBLICENSE-L message of February 12, 1997, enumerated a dozen different pricing models for electronic resources, and correspondents added several more in subsequent discussion.

15. Several reasons are advanced for the higher cost of electronic resources over comparable print resources: (1) the producers are making new R&D and technology investments whose significant prices are passed on to the customer; (2) producers of journals generally offer a package which includes print plus electronic versions, giving the customer two different forms of the same information, rather than one only; (3) the functionality of e-resource is arguably higher than of the print version; (4) electronic resources are not marketed as single journals or books but as scaled-up collections, often of substantial heft (consider the corpora of humanities full texts marketed by Chadwyck-Healey, the large backfile collections of JSTOR, the full collection of Academic Press titles available under its IDEAL program: it seems that there is little incentive for producers to create and sell one electronic item at a time); and (5) in becoming the source or site or provider, the electronic information is taking on many of the library's roles and costs.

16. A LIBLICENSE-L message of March 14, 1997, defined aggregators in the following way: "Aggregation" as used on this list means the bundling together or gathering together of electronic information into electronic collections that are marketed as a package. For example, DIALOG@CARL aggregates 300 databases; Academic Press's IDEAL aggregates 170+ journals; Johns Hopkins's Project MUSE is an electronic collection of 40+ journals, and so on. But the term "aggregator" is more usually used in describing the supplier who assembles the offerings of more than one publisher, so one is more likely to hear Dialog, OCLC, Information Access, and UMI spoken of as aggregators, than The Johns Hopkins University Press.

17. License negotiations between libraries and producers now do take into account the matter of electronic archiving or at least the parties pay lip service to perpetual access. For example, it is common for an electronic resource license to offer some form of access or data if a the library cancels a license or if the provider goes out of business. However, while the license addresses this matter, the underlying solutions are far from satisfactory for either party. We leave the matter of archiving, a huge topic and concern, to other papers and venues; clearly the whole underpinnings of libraries and culture are at stake depending on the outcomes of the archiving dialogs that are in place now and will surely outlast our lifetimes.

18. At Yale, for example, after close discussions on this matter with the Library to make sure that points of view were in synch, General Counsel delegated library content licensing to senior library Administration and it is now done by the Associate University Librarian for Collections, with considerable support and backstopping by Yale's public services and collections librarians in effective and productive teamwork.

19. In fact, the software development was funded by the Council, now known as CLIR, in June 1997 and its product should be available on the WWW site by year end.

20. The case *Princeton University Press v. Michigan Document Services, INC.*, asked the question: does a copy shop infringe on publishers' copyrights when it photocopies coursepack materials? This material comprises book chapters and articles for students of nearby colleges and universities. The owner of MDS argued that he was copying on behalf of the students and exercising their fair use rights. The recent ruling on appeal in the Sixth Circuit was for the publishers. For extensive documentation on this matter, see Stanford's Fair Use site: http://fairuse.stanford.edu/mds/

21. For the journals available through Stanford's HighWire, see: http://highwire.stanford.edu.

Scholarly Communication and Technology

Conference Organized by The Andrew W. Mellon Foundation

at Emory University
April 24-25, 1997

Copyright © of the papers on this site are held by the individual authors or The Andrew W. Mellon Foundation. Permission is granted to reproduce and distribute copies of these works for nonprofit educational or library purposes, provided that the author, source, and copyright notice are included on each copy. For commercial use, please contact **Richard Ekman** at the The Andrew W. Mellon Foundation.

Session #7 Multi-Institutional Cooperation

The Cross Currents of Technology Transfer: The Czech and Slovak Library Information Network.

Andrew Lass
Project Manager, Czech and Slovak Library Information Network
Mount Holyoke College

. . . .

Andrew Lass
Department of Sociology and Anthropology
Mount Holyoke College
© June, 1997

I. Introduction

One would have no great difficulty in estimating the demand function, i.e., the relationship between the price and the quantity that can be sold at that price for, say, tomatoes. But one would have considerable problems in making sales predictions at

> various hypothetical alternative prices for a new product that looks like a blue tomato and tastes like a peach. (Quandt 1995:20)

This vivid image of an odd looking vegetable that tastes like a fruit is meant to highlight the difficulty of estimating the demand side in the overall cost picture of producing and distributing new products, such as electronic publications. Compared to the 'traditional' printed material, electronic products *are* new, from their internal architecture to the mechanisms of production, distribution, and access that stem from it. As the author of the above quote points out: "Econometric approaches are well developed for estimating the demand for given products, but face greater difficulties in estimating the demand for products with as yet untested combinations of characteristics" (ibid.). After all, the world of readers is not a homogeneous social group, a market with a simple set of specific needs. Yet, we assume that a segment of this market--the scholarly community--takes easily and *more or less* quickly to supplementing their long established habits (of associating the printed text with a paper object) with different habits, experienced as equally convenient, of searching for and reading electronic texts. While this may be so, it should be emphasized *at this point* that it is precisely in the expression "more or less" that the opportunity lies--for those of us interested in transitions--to see what is involved in this change of habit and why it is not just a "matter of time." As anyone who has tried to explain the possibilities of electronic text delivery to an educated friend will attest, the idea is viewed with anxiety, it is taken to mean the end of the book. The Minister of Culture of the Czech Republic, a well known author and dissident, looked at me with surprise as I tried to explain the need for library automation (and therefore for his ministerial support); he held both hands clasped together as if in prayer and then opened them up like a book close to his face. He took a deep breath, exhaled and explained how much the scent of books meant to him. A rather daring leap of the imagination--from on-line cataloguing and microfilm preservation to the demise of his personal library--but not an uncommon one, even among those who should know better. What was I to say to a professor of aesthetics at Charles University in Prague who demanded to know the truth about "that library project you are involved in" at the National Library in Prague? She had gone to pick up a book ordered through ILL and was advised by the person attending the circulation desk that she had better photocopy it because, "once they install the computers we will stop lending the books." It is not just the community of scholars, then, but librarians and politicians who must change their attitudes and habits. The problem is further compounded, and the blue tomato cum peach extended, if we consider that in the case of Eastern Europe this new product is being introduced into a setting where the very notion of a market is itself unsettled. The question of demand is quite different in a society that had been dominated by a political economy of command.

Needless to say, these humorous examples are neither particular to Eastern Europe nor to information technology (fantastic expectations have accompanied the introduction of many innovations that have changed the way we live). They merely highlight the period of transition when a blue tomato tastes like a peach. The important point about these perceptions is that, like all perceptions, they reflect a world of expectations "already in place" (what anthropologists would call culture) and they inform actions that, intended to change that world also end up reinforcing it. It is no different with information technology and its relation to scholarly research.

In the pages that follow I will give a simplified account of an extensive library automation and networking project that the Andrew W. Mellon Foundation initiated and funded abroad, in the Czech and Slovak republics. My aim is critical rather than comprehensive. By telling the reader about some of the obstacles that have been confronted along the way, I hope to draw attention to the kinds of issues that need to be kept in mind when we think of establishing library consortia--the seemingly natural setting for the new technologies--in other countries. The story told here is but part of the whole picture. But it is an essential part. In so far as it is about "transition," it is also about the kinds of things that take place before other things can follow.

II. *The CASLIN projects*

The Mellon-funded proposal to establish *The Czech and Slovak Library Information Network* (CASLIN) commenced in January 1993. In its original stage it involved four libraries in what has now become two countries: the National Library of the Czech Republic (in Prague) and the Moravian Regional Library (Brno) the Slovak National Library (Martin) and the University Library of Bratislava. These four libraries had signed an agreement (a "Letter of Intent") that they would cooperate in all matters that pertained to fully automating their technical services and, eventually, in developing and maintaining a single on-line Union Catalogue. They also committed themselves to introducing and upholding formats and rules that would enable a "seamless" integration into the growing international library community. For example, compliance with the UNIMARC format was crucial in choosing the library system vendor (the bid went to ExLibris's ALEPH). Similarly, Anglo-American cataloguing rules (AACR2) have been introduced and, most recently, there is discussion of adopting the LC subject headings. Needless to say, the implementation was difficult and the fine tuning of the system is not over yet, though most if not all of the modules are up and running in all four libraries. The first on-line OPAC terminals were made available to readers during 1996. At present, these electronic catalogues reflect only the library's own collection--there are no links to the other libraries, let alone to a CASLIN Union Catalogue--though they do contain a variety of other databases (for example a periodicals distribution list is available on the National Library OPAC which lists the location of journals and periodicals in different libraries in Prague, including the years and numbers held). A record includes the call number--a point of no small significance--but does not indicate the loan status, nor does the system allow someone to 'Get' or 'Renew' a book.[1] In spite of this, the number of users of these terminals has grown sharply, especially among university students, and librarians are looking for ways to finance more (including some graphic ones with access to the WWW).

In the period between 1994 and 1996 several additional projects (conceived as extensions of the original CASLIN project) were presented to the Mellon Foundation for funding. It was agreed that the new partners would adopt the same cataloguing rules, as well as any other standards, and that they would (eventually) participate in the CASLIN Union Catalogue. The idea was to help assure that the original 'backbone' got some 'arms and legs' and would thereby have a more lasting impact on future library trends in both countries. Our vision was to come closer to a setup in which bibliographic as well as other types of information would be more easily accessible throughout the region and, considering the libraries' decreased purchasing powers, at lower costs. Each one of these projects also posed a unique challenge to the use of information technology as an integrator of disparate and incongruous institutional settings.

The Library Information Network of the Czech Academy of Science (LINCA) was projected as a two-tiered effort that would a) introduce library automation to the central library of the Czech Academy of Sciences and thereby b) set the stage for the building of an integrated library-information network that would connect the specialized libraries of all the 60 scientific institutes into a single web with the central library as their 'hub.' At the time of this writing, the central library's LAN has been completed and most of the hardware installed, including the high capacity CD-ROM (UltraNet) server. The ideal of connecting all the institutes will be tested against reality this year, when the modular library system (BIBIS by Square Co., Holland) will be introduced together with workstations and/or mini-servers in the many locations in and outside the city of Prague. This is the first of the Mellon-funded CASLIN projects designed specifically with the idea in mind of integrating 'traditional' library functions with electronic text delivery, in particular, the availability of CD-ROM databases. Unlike the libraries in the original CASLIN project, all of which are defined as 'public' libraries (though their collections are large, historically valuable, and specialized), the libraries of the Academy of Sciences are meant to cater to the needs of primary research.[2]

The Kosice Library Information Network (KOLIN) is an attempt to draw together three different institutions (two universities and one research library) into a single library consortium. If successful, this consortium in eastern Slovakia would comprise the largest on-line university and research library group in that country. The challenge lies in the fact that the two different types of institutions come under two different government oversight ministries (of Education and of Culture) which further complicates the already strained budgetary and legislative setup. Furthermore, one of the universities--the University of Pavel Josef Safarik (UPJS)--at that time had two campuses (in two cities 40 km apart) and its libraries dispersed among 13 locations. UPJS is also the Slovak partner in the Slovak-Hungarian CD-ROM network (Mellon-funded HUSLONET) that shares in the usage and the costs of purchasing database licenses.[3]

Finally, the last of the CASLIN "add-ons" involves an attempt to bridge incompatibilities between two established library software systems by linking two university and two state scientific libraries in two cities (Brno and Olomouc) into a single "regional" network, *The Moravian Library Information Network* (MOLIN). The two universities--Masaryk University in Brno and Palacky University in Olomouc--have already completed their university-wide library network with TinLib (of UK) as their system of choice. Since TinLib records do not recognize the MARC structure (the CASLIN standard adopted by the two state scientific libraries) a "conversion engine" has been developed to guarantee full import and export of bibliographic records. Though it is too soon to know how well the solution will actually work, it is clear already that its usefulness goes beyond MOLIN, since TinLib has been installed in many Czech universities.[4]

In order for any of these projects to make sense other key document delivery functions would have to be taken care of. Fortunately, storage, document preservation, retrospective conversion and connectivity have all undergone substantial changes over the past few years. They are, however, worth a brief comment since they add important background information to the theme of this paper.

1. Access to holdings was limited by the poor condition of the physical plant and, in the case of special collections, the actual poor condition of the documents. The National Library in Prague was the most striking example of this situation; it was in a state of de facto paralysis when I first contacted the institution in 1990. Of its close to 4 million volumes only a small percentage was accessible. The rest were literally "out of reach" because they were either in milk crates and unshelved, or in poorly maintained depositories in different locations around the country.[5] This critical situation turned the corner in January, 1996, when the new book depository of the NL was officially opened in the Prague suburb of Hostivar. Designed by the Hillier Group (Princeton, N.J.) and built by a Czech contractor, it is meant to house 4.5 million volumes and contains a rare book preservation department (including chemical labs) and a large microfilm

department. As a result of cleaning, moving and reshelving over 2 million volumes by the end of 1996, it is now possible to receive the books ordered at the main building (a book shuttle guarantees overnight delivery).[6] Other library construction has been under way, or is planned, for other major scientific and university libraries in the Czech Republic.[7] But, of course, there is more to this proliferation of building projects. Objectively speaking, it is true that there was little if any attention paid to libraries (let alone to the pressing storage needs) during over half a century, so unless this matter was dealt with in a serious way, university and research libraries could not fulfill their tasks. On the other hand, there is also a significant component of symbolic value and political prestige involved. There is nothing unusual in this or in the tension it provokes or fuels. But most of the building plans incorporate the idea of computer automation and Internet access; they do not take into consideration what impact the possibility of a "virtual library" may have on their design and cost estimates. This is in spite of the progressively worse budgetary constraints, and in spite of the fact that information on this is readily available and many of the librarians had attended seminars on this topic. We will see this lack of vision repeated in other contexts.

2. The original CASLIN project included a small investment in microfilm preservation equipment, including a couple of high-end cameras (GRATEK) with specialized book cradles--one for each of the National Libraries--as well as developers, reader-printers and densitometers. The idea was to a) preserve the rare collection of 19th and 20th century periodicals (that are turning to dust), b) significantly increase the turnaround time that it takes to process a microfilm request (from several weeks to a few days), and c) make it technically possible to meet the highest international standards in microfilm preservation and consequently guarantee digital scanning and the conversion to other media in the future.[8]

3. The most technologically ambitious undertaking, and one that also has the most immediate and direct impact on document accessibility, is the project for the retrospective conversion of the general catalogue of the National Library in Prague. Known under the acronym RETROCON, it involves a laboratory-like setup of hardware and software (covered by a Mellon Foundation grant) that would--in a virtually assembly-line fashion--convert the card catalogue into Aleph-ready electronic form (UNIMARC). It is designed around the idea of using a sophisticated OCR in combination with a specially designed software that semi-automatically breaks down the converted ASCII record into logical segments and places them into the appropriate MARC field. This software, developed by a Czech company (COMDAT) in cooperation with the National Library, operates in a Windows environment and allows the librarian to focus on the "editing" of the converted record (using a mouse and keyboard, if necessary), instead of laboriously typing in the whole record. As an added benefit, the complete scanned catalogue has now been made available for limited searching (under author and title in a Windows environment), thereby replacing the original card catalogue.[9] Logically, the bibliography of 20th century Czech publications has been the top priority. It is the most used and its "automatic" conversion is thought to be the least problematic, since most of the records already exist in print and follow a standardized format (i.e., it is most amenable to an OCR to UNIMARC conversion algorithm). One of the most interesting aspects of this project has been the "out-sourcing" of the final step in the conversion to other libraries, a sort of division of labor (funded in part by the Ministry of Culture) that increases the pool of available 'expert' cataloguers. In exchange for the work, libraries get to keep the basic equipment and, as a side of effect of working with the COMDAT software, they also learn to catalogue in the UNIMARC record structure.[10]

4. For the most part, all installations of the LAN have proceeded with minimal problems and the library-automation projects, especially as these involve the technical services, are finally up and running. A quite different story can be told about the larger framework. While the Czech Intern group (CESNET) has been around for a while, its throughput had been very low (the first 10MB links were offered in January, 1997). This has had an adverse effect on library management, especially of the CASLIN consortium as a whole. The irony of the situation is made apparent by the fact that it is possible, and has been for several years, to log on to the catalog of the CASLIN libraries via Telnet on a home computer from abroad. The phone system in the Czech and Slovak republics has yet to undergo a serious overhaul (the first digital switchboards are now being installed) and even then, connecting to the Net from, for example, a Prague home, is very slow. Since all local calls are toll calls, it is only for members of the new entrepreneurial class, who can afford it, and for the growing number of computer addicts. Up until now, logging into the National Library from my friend's place across the street has been technically difficult and quite expensive.

III. *Cross-currents*

If one compares the present condition and the on-line readiness of research and university libraries in Central Europe with the *status quo*, as it arrived at the doorstep of the post 1989 era, then there can be no doubt that dramatic improvements have taken place. But if the once empty (if not broken) glass is now half filled, it also remains half empty. Certainly, that is how most of the participants tend to see the situation. Maybe because they are too close to it and because chronic dissatisfaction is a common attitude. Yet the fact remains that throughout the implementation and in all of the projects, obstacles appeared at just about every step of the way. While most of them were resolved, though not without some cost, all of them can be traced basically to three sources of friction: a) those best attributed to "external" constraints--the budgetary, legal, political and, for the most part, bureaucratic ties that directly affect a library's ability to function and implement change, b) those caused by "cultural misunderstandings"--the different habits, values and expectations that inform the activity of "localization," and c) the "internal" problems of the libraries themselves, no doubt the most important locus of micro-political frictions, and therefore of problems and delays. In what follows, I will focus on the first (with some attention paid to the second), since my emphasis here is on the changing relations *between* what are taken to be separate institutional domains (particularly between libraries and other government organizations or the market) as I try to make sense of the persistently problematic relationships between libraries (particularly within the CASLIN group). Obviously, while these analytical distinctions are heuristically valuable in reality these sources of friction are intertwined and further complicated by the fact that the two countries are undergoing a transition full of aftershocks and endless series of corrections. It is not only the libraries that are being transformed, so is the world of which they form a part.[11] To make sense of this double transition and to describe the multifaceted process that the library projects have moved through may pose some difficulties. But it also offers a unique opportunity to observe whether, and, if so, how, the friction points move over time. What could have been predicted when the initial project commenced--that "implementation" and "system localization" would also mean giving in to a variety of constraints--is only beginning to take on the hard contours of reality four years later. In several instances the results differ from *our* initial conception, but I don't think it would be fair to assume that the final outcome will be a compromise. Instead, the success of the Mellon library projects in Eastern Europe (of which CASLIN is only one) should be judged by the extent to which they have been accepted and have taken on a life of their own, initially distinguishable but finally inseparable from the library traditions already in place. After all, if the projects were designed to affect a change in the library system, and by system we must understand a complex of organizational structures, a real culture, and an actually existing social network, then we must also expect that it will respond that way, as a complex socio-cultural system. What appeared at first as a series of stages (goals) that were to follow one another in logical progression and in a "reasonable" amount of time may still turn out to have been the right series. It's just that the progression will have followed another (cultural) logic, one in which other players--individuals and the organizational rules that they play by--must have their part. As a result, the time it actually takes to get things done seems "unreasonable," and some things even appear to have failed because they have not taken place *as* and *when* expected. What does this mean? A seemingly philosophical issue takes on a very real quality as we wonder, for example, about the future of the CASLIN consortium. If establishing a network of library consortia was one of the central aims of the Mellon project, then it is precisely this goal that we have failed to reach, at least now, when it was supposed to be long in place according to our scheme of things. There is no legal body and no formal association of participating libraries in place, This is particularly important and, needless to say, frustrating for those of us who take for granted the central role that networking and institutional cooperation play in education and scholarly research. But behind this frustration another one hides: it is probably impossible to say whether what is experienced as the status quo, which in this case is perceived as a failure or shortcoming, is not just another unexpected curve in a process that follows an uncharted trajectory.[12]

As I have noted above, in 1992 a "Letter of Intent" had been signed by the four founding CASLIN members. It was a principal condition of the project proposal. In January 1996, when this part of the project was--for all intents and purposes--brought to a close, there was still no formally established and registered CASLIN association with a statute, membership rules and a governing body in place. Although the four libraries had initially worked together to choose the HW and SW, the work groups that had been formed to decide on specific standards (such as cataloguing rules, language localization or the structure of the union catalogue record) had trouble cooperating and their members often lacked the authority to represent their institution. If things got done, it was due more to the enthusiasm of individuals and the friendly relations that developed among them than because of a planned, concerted effort on the part of the library leadership guided by a shared vision. If anything, there was a sense, at times, that the prestige of the project was more important than its execution or, more exactly, that while the funding for library automation was more than welcome, so was the political capital that came with being associated with this U.S.-funded project *even if this meant using the capital at a cost to the consortium.* As is well documented from many examples of outside assistance in economic development, well intentioned technology transfer is a prime target for subversion by other, local intentions; it can be transformed with ease into a pawn in another party's game. Potential rivalries and long standing animosities that existed among some of the libraries, instead of being bridged by the project, seemed to be exacerbated by it. In one instance, for example, affiliation with the Mellon project was used by a library to gain attention of high government officials (such as the cultural minister) responsible for policies affecting their funding and, most importantly, their mandate. The aim, as it now turns out, was to gain the status of a national library. The target, the library that already had this status, the Slovak National Library, was its primary CASLIN partner. While both libraries participated in the CASLIN project's implementation, and even cooperated in crucial ways at the technical level (as agreed), their future library cooperation was being undermined by a parallel, semi-clandestine, political plot. Needless to say, this has left the CASLIN partnership weakened and the library managements disfunctional.[13]

As the additional library projects, mentioned earlier, were funded and the new libraries joined the original CASLIN group, it became clear that the new, larger group existed more in rhetoric than in fact. From the newcomer's point of view there was not much "there" to join. "What was in this for us, and at what cost?" seemed to be the crucial question at the January 1996 meeting at which a written proposal for a CASLIN association was introduced by the National Library in Prague. This was not the first time that an initiative had been presented only to fail to take hold. Nor was it the last. The discussion about the proposal resulted in a squabble. An e-mail discussion group was established to continue the discussion, but nothing came of that either.

If the point of a consortium is for libraries to cooperate in order to benefit (individually) from the sharing of resources so as to provide better on-line service, then a situation such as this one must be considered counterproductive. A year later (January, 1997) a meeting was arranged for all the CASLIN members and other university and scientific libraries were once again invited to attend. This time no official proposal to establish an association was put forward and, instead, progress reports were given on specific aspects of library automation. Since these came from the National Libraries, which are mandated to develop and maintain (national) standards, they were of immediate interest to all the attending libraries. The detailed reports on the retrospective conversion project, and on the development of the CASLIN union catalogue record standard, made it clear that some cooperation was continuing at the practical level of technical services. But in the discussion that followed several library directors, many of whom were not CASLIN members (but who were clearly interested in such a possibility), expressed concern that, without more cooperation at all levels, it was going to become more difficult for individual libraries to participate in on-line consortia. Between budgetary problems, lack of expertise and unpredictable vendors (and their ever-changing products and standards) the smaller libraries, in the words of one of the librarians, "will be left out in the dark where they are liable to make costly mistakes that could have been avoided." This call for help presents a glimmer of hope, at least in the sense that "an expression of need," as one library director put it to me, "is what it will take for an association to form." Nothing came out of this meeting either.

How does one explain CASLIN's chronic inability to get off the ground as a real existing organization? The sense of apathy, reluctance, or even antagonism: where does it come from? For one, the fact that all the original CASLIN libraries come under the administrative oversight of the Ministry of Culture goes a long way in explaining the persistence of territorial behavior. The dramatic cuts in the ministries' overall budgets are passed down to the beneficiaries who find themselves competing for limited goods. If the difference from the previous setup (under the "planned" socialist economy) lies with the fact that the library has the status of a legal subject that designs and presents its own budget, its relationship to the Ministry--very tense and marked by victimization--seems more like the "same old thing." In other words, certain aspects of organizational behavior continue not only by force of habit (a not insignificant factor in itself), but also because these are reinforced by a continuing culture of co-dependency and increased pressure to compete over a single source of attention. As if, from our point of view, the formal command economy has been transformed into a market economy only to the extent that strategic and self-serving positioning is now more obvious and potentially more disruptive. So called "healthy competition" (so called by those whose voices dominate in the present government and who believe in the self-regulating spirit of "free market forces") seems to show only its ugly side; we see the Mellon project embraced with eagerness *in part* because of the way its prestige could be used to gain a competitive advantage over other libraries. In the case of CASLIN partners, we see it take the form of suspicion, envy and even badmouthing expressed directly to the Mellon grants administrator (myself).[14]

What the are the constraints under which a research or national library operates, and in what way is the present situation different from the "socialist" era [1948-1989]? An answer to these questions will give us a better sense of the circumstances under which attempts to bring these institutions up to international standards--and get them to actively cooperate--must unfold. Figures 1. and 2. illustrate *the external ties between a library and other important domains of society that affect its functioning and co-define its purpose* before and after 1989 (while keeping in mind that economic, legal and regulatory conditions have been in something of a flux in the years since 1989 and, therefore, that the rules under which a library operates continue to change).

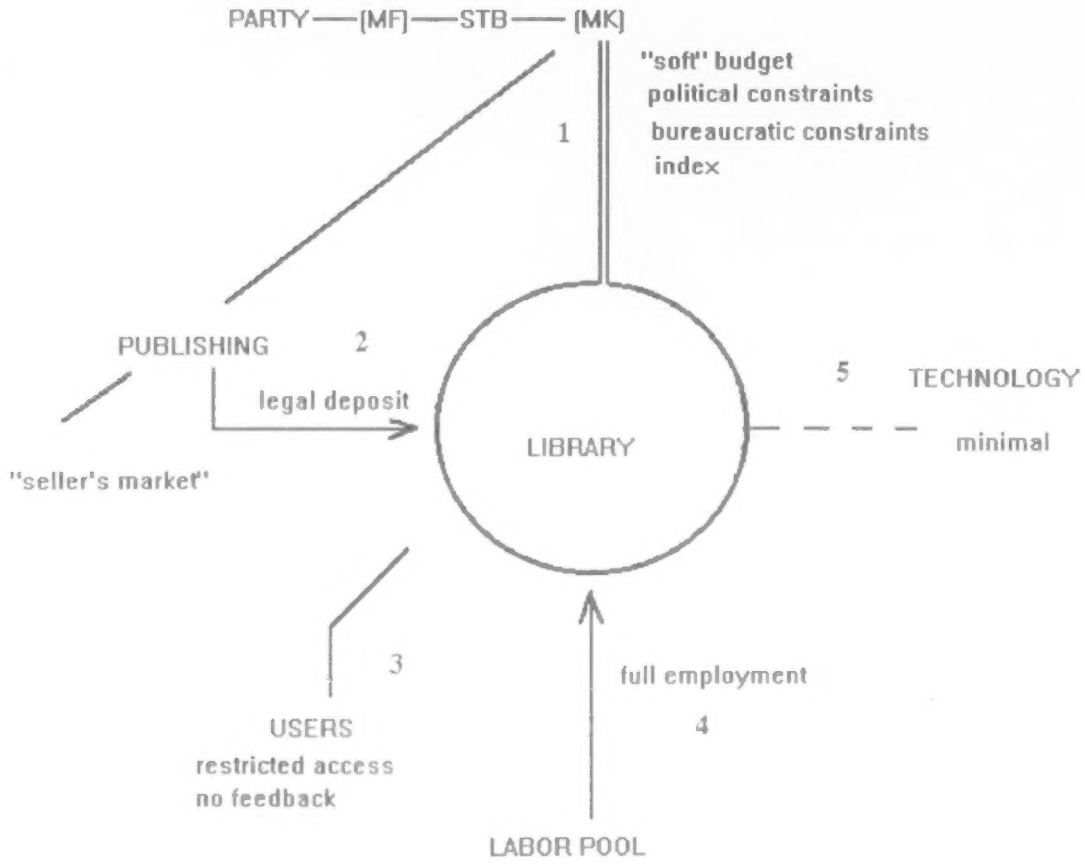

PARTY—[MF]—STB—[MK]
"soft" budget
political constraints
bureaucratic constraints
index
1
PUBLISHING
2
legal deposit
"seller's market"
LIBRARY
5
TECHNOLOGY
minimal
3
USERS
restricted access
no feedback
full employment
4
LABOR POOL

Figure 1: Czech research library *before* 1990: external ties

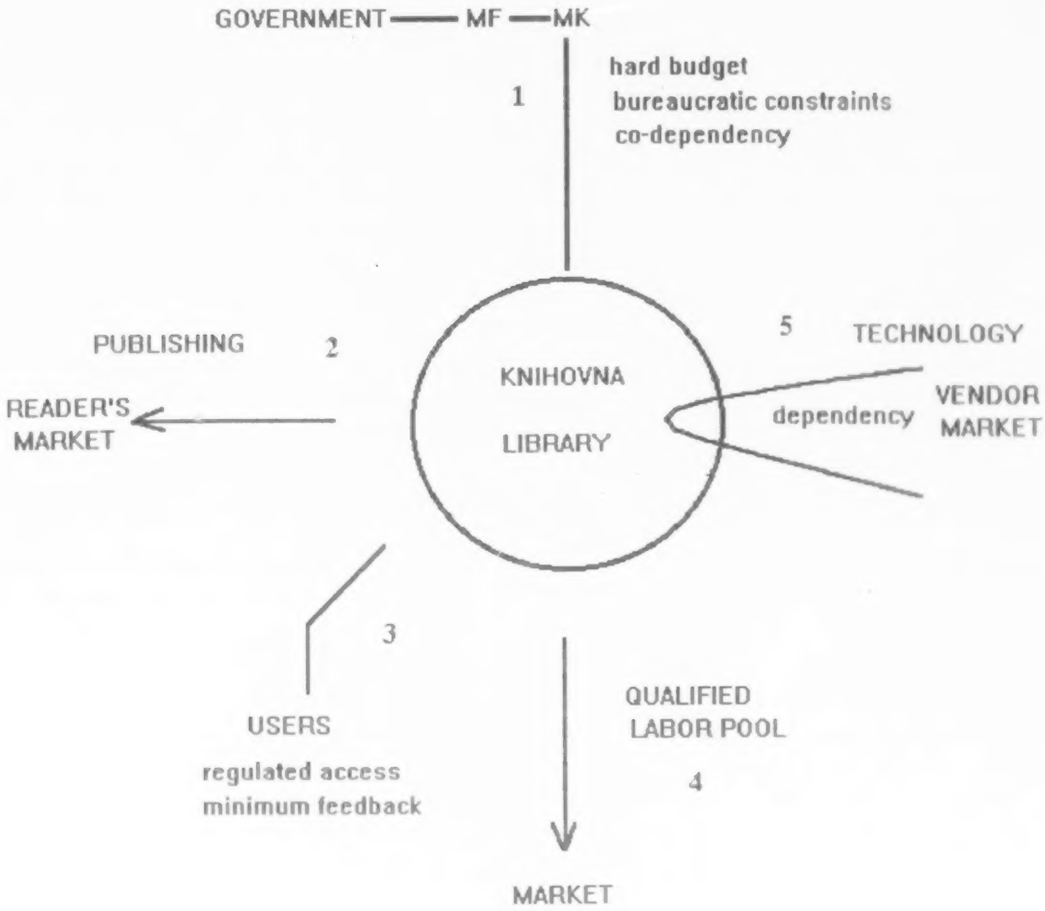

GOVERNMENT — MF — MK
1
hard budget
bureaucratic constraints
co-dependency
PUBLISHING
2
READER'S MARKET
KNIHOVNA
LIBRARY
5
TECHNOLOGY
VENDOR MARKET
dependency
3
USERS
regulated access
minimum feedback
QUALIFIED LABOR POOL
4
MARKET

Figure 2: Czech research library *after* 1990: external ties

1. Under "party" rule, the library, like all other organizations, came under direct control of its ministry, in this case the Ministry of Culture [MK]. One could even say, by comparison with the present situation, that the library was an extension of the ministry. However, the ministry was itself an extension of the centralized political rule (the Communist party), including the watchful eye of the secret police [STB]. The director was appointed "from above" [PARTY] and the budget "arrived" from there as well. While requests for funding were entertained, it was hard to tell what would be funded and under what ideological disguise.[15] For the most part, the library was funded "just in order to keep it alive," though if the institution ran out of money in any fiscal year, more could be secured to "bail it out" (hence the expression "soft budget"). In addition to many bureaucratic constraints (regarding job descriptions and corresponding wage tables, building maintenance and repairs or the purchase of monographs and periodicals), *many of which remain in place*, there were political directives regarding "employability"[16] and, of course, the ever-changing and continuously growing list of prohibited materials to which access was to be denied (Index). In contrast, the library is now an "independent" legal body that can more or less decide on its priorities and is free to establish working relationships with other

(including foreign) organizations. The decision making, including organizational changes, now resides within the library. While the budget is presented to the ministry and is public knowledge, it is also a hard budget that is, in the end, set at the ministerial level as it matches its cultural policies against those of the Ministry of Finance [MF] (and therefore of the ruling government coalition). After an initial surge in funds (all marked for capital investment only), the annual budgets of the libraries have been cut consistently over the past 5 years (i.e., they are not even adjusted for inflation but each year are actually lower than the previous one). This has seriously affected the ability of the libraries to carry out their essential functions, let alone purchase documents or be in the position to hire qualified personnel.[17] For this reason, I prefer to speak of a relationship of "co-dependence." Since the Ministry of Culture still maintains direct control over the library's ability to actualize its "independence"--though it has gradually shifted from an attitude of outright harassment to one of more genuine interest--I remain skeptical that, should these ties be further weakened, if not cut, either of the institutions would know how to manage without the other. The point is, that where the Ministry of Culture is supposed to oversee the well-being of the institutions it oversees, it is, as is usually the case in situations of government supervision, perceived as the powerful enemy.

2. The publishing world was strictly regulated under the previous regime: all publishing houses were state enterprises (any other attempt at publishing was punishable by law) and all materials had to pass the scrutiny of the state (political) censor. Not everything that was published was necessarily "political trash" and editions were limited; the resulting economy of shortage created a high demand for printed material, particularly modern fiction, translations from foreign languages and the literary weekly (hence "sellers market"). Libraries benefited from this situation. Because all state scientific and research libraries were recipients of the legal deposit, their (domestic) acquisitions were, de facto, guaranteed. At present, the number of libraries covered by the deposit requirement has been reduced from some three dozen to half a dozen. This change was meant to ease the burden on publishers and give the libraries a freer hand in building their collection in a "competitive marketplace." But considering the severe cuts in the budget, many of the libraries cannot begin to fulfill even the most spartan acquisitions policy. For the same reason publishers, of whom there are many and all of whom are private and competing for the readers' attention, do not consider libraries as important parts of their market. Furthermore, many of the small and often short-lived houses do not bother to apply for the ISBN or to send at least one copy (the legal deposit law is impossible to enforce) to the National Library which, in turn, cannot fulfill its mandate of maintaining the national bibliographic record.

3. During the Communist era, access to materials was limited for several obvious reasons: political control (books on the "index," the limited number of books from Western countries, and theft) or deliberate neglect (the progressively deteriorating storage conditions eventually made it impossible to retrieve materials). Over the years, in effect, there was less and less correspondence between the card catalogues in the circulation room and the actual holdings. As a result, for example, students and scholars stopped using the National Library in Prague because it was increasingly unlikely that their requests would be filled. This was also true for recent Czech or Slovak publications because of an incredible backlog in cataloguing or because the books remained unshelved. Of course, in such a system there was no place for user feedback. Since then, some notable improvements, many of them due to Mellon and other initiatives, have been made in public services such as self-service photo-copying machines and, to remain with the example of the National Library, quick retrieval of those volumes that have been reshelved in the new depository. Also, readers are now used to searching the electronic OPACs or using the CD-ROM databases in the reference room. On the other hand, the backlog

of uncatalogued books is said to be worse than before and, with acquisitions cut back and the legal deposit not observed, the reader continues to leave the circulation desk empty handed. The paradoxical situation is not lost on the reader: if the books are out of print or, as is more often the case these days, their price beyond what they could afford, going to the library may not be a solution either. So far the basic library philosophy has remained the same as it has throughout its history: while there is concern for the user, libraries are not genuinely "user driven" (only a few university libraries have adopted an open stack policy) and, as far as I can tell, user feedback is not a key source of information, actively sought and used in setting priorities.

4. Under the policies of socialist economy, full employment was as much a characteristic of the library as it was for the rest of society. In effect, organizations "hoarded" labor (as they did everything else) with a sort of "just in case" philosophy in mind, since the point was to fulfill "the plan," at just about any cost, and provide full benefits for all, with little incentive for career development (other than through political advancement). Goods and services got to be known for their poor quality, the labor force for its extremely low productivity and its lousy work morale. More time seemed to be spent in learning how to "trick the system" than in working with it, to the point where micro-political intrigue--the backbone of the "second" economy--competed very well with the official chain of command. The introduction of a market economy after 1990 did very little to help change this in a library, a state organization with no prestige. Simply put, the novelty and promise of the private sector, coupled by its high employment rate and good wages, has literally cut the library out of the competitive market for qualified labor. Between the budget cuts and the wage tables still in place there is little space left for the new management to negotiate contracts that would attract and keep talented people in the library. Certainly not those with an interest in information technologies and data management.[18]

5. As mentioned above, the first information technologies arrived in the state scientific and national libraries in the late 1980s. Their impact on budgets was minimal (Unesco's ISIS is free ware), as was their effect on technical services. On the other hand, the introduction of information technologies into these libraries, in particular the CASLIN group, was the single most visible *and disruptive* change, a sort of wedge that split the library organizations open, that has occurred since 1990 (or, according to some, during the last century). The dust has not yet settled, but, in view of our present discussion, one thing is clear already: between the Mellon funds and the initial capital investment that followed, libraries have become a significant market for the local distributors of hardware and for the library software vendors (in contrast to the relationship with publishers). But, as everywhere else in the world of information technologies, these are not one-time purchases, but only the first investments into a new kind of dependency, a new "external" tie that the library must continue to support and at no small cost. And not just financial cost. The ongoing complications with the technology and the chronic delays in systems localization only contribute to the present sluggish state of affairs, and thus lend support to the ever cynical factions within the organization that knew "all along" that "the whole automation project was a mistake." Obviously, the inability to attract qualified professionals doesn't help.

What I have painted here is but part of the picture (a detailed analysis of the micropolitics that actually go on, both inside the organization and in relation to other organizations, particularly other libraries, would make up the other part). But the above discussion should help us see how and why the libraries feel trapped in a vicious circle from which they perceive little or no way out, other than continuing to battle for their place in the sun. Of course, their tactics and battle cries only reinforce the relationship of codependency, as well as their internal organizational problems. And, "from the outside," that is exactly what the public and government officials see:

that these are institutions that need to "grow up" and learn what "real work" is before "more money is poured down the drain." Needless to say, I don't think there is any doubt that a government that has made a conscious choice against long term investment into the educational, scientific and information sectors must carry a sizable portion of the blame.

If the long-standing administrative ties between libraries and the Ministry of Culture inform and override the building of new, potentially powerful ties to other libraries, then the flip side of this codependency, its result, is *a lack of experience with building, and envisioning the practical outcome of a horizontally integrated (i.e., non-hierarchical) association of independent organizations.* The libraries had had only limited exposure to automation, and the importance of long-term strategic planning was lost on some of them.[19] At least two other factors further reinforced this situation: the slow progress (the notorious delays mentioned above) in the implementation of the new system, which had involved what seemed like impractical and costly steps (such as working in UNIMARC), and the sluggish Internet connection. This suggests that *at the present, a traditional understanding of basic library needs (which themselves are overwhelming) tends to take precedent over scenarios that appear much too radical and as not grounded in a familiar reality. Since the on-line potential is not fully actualized, its impact is hard to imagine and so the running of the organization in related areas continues to be predominantly reactive rather than proactive. In other words, "in house" needs are not related to network solutions.* Especially when such solutions appear to be counter intuitive for the established (and more competitive) relationship between the libraries!

Cooperation among the libraries existed at the level of system librarians and other technical experts. Without this cooperation the system would not have been installed, and certainly not as an identical system in all four libraries. In addition (and, I should say, ironically), the CASLIN project has now received enough publicity to make it a household name among librarians. The acronym has a life of its own, and there is a growing interest among other scientific libraries to join this "prestigious" group (that both does and does not exist).[20] The meetings described above were witness to the moment at which the confluence of *de facto* advances in technical services and a growing interest of other libraries in logistical support (involving technology and technical services) created a palpable *need* for a social organization that would exist a) above and beyond the informal network of cooperation and b) without association with the name and funds of the Andrew W. Mellon Foundation (its original reason for existence). I have heard it said, on the other hand, that "nothing more is needed," since the fundamentals of CASLIN are now embedded in the library process itself (reference here, I gather, was to cataloguing), and in the existing agreements between individual libraries on the importing and exporting of records into and from the CASLIN Union Catalogue which is serviced by the two National libraries. In fact, as the most recent meeting (June 1997) of the Union Catalogue group made clear, this is indeed where the seed of an association of CASLIN libraries lies. The import and export of records and the beginning of the UC database have yet to materialize, but that was what brought these individuals who represented individual libraries together. If only they had the experience and the wherewithal to run their own show and stick to it. But if they do, then there is fair chance that an organization of CASLIN libraries will take off.

IV. Concluding remarks

The above discussion raises three very important points for me as I begin to gather my thoughts on the Mellon CASLIN projects. *First*, regarding *cultural misunderstanding*, the problem with the "misbehaving consortium" may lie to a large extent with our (e.g., U.S.) expectations of what cooperation looks like and what basic fundamentals an on-line library consortium must embrace in order to do its job well. In the Czech and Slovak case, not only were the conditions not in place, they were counter-indicative. While our naivité caused no harm (the opposite is the case, I am repeatedly told!), it remains to be seen what the final result will look like. And that is where the really intriguing lesson resides: maybe it is not so much that we should have or even could have thought differently, and therefore ended up doing "that" rather than "this." Perhaps it is in the (information) technology itself--in its very architecture--that the source of our (mis)understanding lies. After all, these technologies were developed in one place and not another. Our library automation systems obviously embody a particular understanding of technical and public services and an organization of work that share the same culture as a whole tradition of other technologies that emphasized speed, volume (just think of the history of railroads or the development of the "American system" of manufacturing) and, finally, access. Every single paper in this conference volume exemplifies and assumes this world. In transferring a technology from one place to another, an implied set of attitudes and habits is being marketed as well. To this possibility my *second* point lends some support: technology transfer involves a time lag, the duration of which is impossible to predict and that is accounted for by a complex series of micro-political adjustments. It is this human factor that transforms the logical progression in the *projected* implementation process into a much less logical but essentially social act. Thanks to this, the whole effort may fail. Without it, the effort cannot exist. Only after certain problems and not others arise will certain solutions and not others seem logical. And since the technology comes from "elsewhere," are not the problems it provokes in a new setting, and the solutions that seem logical to us, just a bit more complicated than we would have liked to believe? It is no secret that much social change is technology driven (hence our conference). It is less clear, ethnographically speaking, what exactly this means, and even less is known about this process when technology travels across cultural boundaries. There is much to be gained from looking carefully at the different points in the difficult process of implementing projects such as CASLIN. Apparently the ripple effect reaches far deeper (inside the institutions) and far wider (the government, the market and the users) than any one would have anticipated. Before it is even delivering fully on its promise, the original Mellon project is demanding changes in library organization and management. Such changes are disruptive, even counterproductive, long before they "settle in." Nevertheless, it is also, and this is my *third* point, eliciting changes in the relations with the "outside." At least on the Czech side--it is important to emphasize that the situation in the two countries has continued to diverge--the Ministry of Culture has taken a keen interest in supporting library automation. On the Slovak side, unfortunately, the Ministry of Culture has played the mostly negative role of paying lip service to, or even actively undercutting, library initiatives. And this is only one of many positions taken by the Slovak government (and even written into law) that are deliberately aimed at controlling intellectual activity (splitting up universities is another).

What, then, is the discernible impact of the technological changes discussed above on the library user and, more specifically, on scholarly communication? In addition to the improvements in public services already mentioned, some of the newer CASLIN library members, such as university libraries and the central library of the Czech Academy of Sciences, are now offering document delivery services in addition to electronic databases on their LAN. These services, based on actual subscriptions (licenses), are continuously threatened, however, by library cuts. Take, for example, the situation at one central university library which has no funds to subscribe to new CD-ROM databases. Because it operates the UltraNet server, it will (depending on the license) provide restricted or wide area LAN service for schools that have been able to secure a CD-ROM license thanks to their own, usually foreign, grants. Since it is mostly the natural sciences and medical schools that have independently funded projects, their use of current on-line journals and databases is, ironically, better covered than the more "traditional" needs of the humanities and social sciences for whom there are no funds to speak of (the equivalent of a few thousand dollars per year for the acquisition of monographs).

Obviously, an evaluation of technology and scholarly communication in Eastern Europe cannot be limited to a discussion of the possible impact that the shift to on-line library consortia may have. As radical as this change is and certainly will continue to be, it is only the first step. Unless we know as much about the other aspects, about past and present trends in the organization and practice of science and higher education, we cannot do this topic full justice. Once again, there are many ways in which the scholarly tradition differs quite sharply from ours. The distinction between universities and scientific institutes (different budgets, different responsibilities), the non-existence of academic tenure, the 100% dependence on state funding, the planning and defining of politically correct research agendas, as well as the political economy of the publishing world, are just some of the features that defined the culture of research and education for 45 years. And only some of these parameters have changed (gradually) since 1990. The significance of these changes is still unclear and a matter of debate. While experimental scientists are long accustomed to the use of the citation index or, for example, chemical abstracts (they may even find their own work cited), the social sciences and especially the humanities operate in a less international context, and are used to their own tradition of accessing, quoting, referencing and even writing. It is difficult to say whether this will or should change, or how this portion of the intelligentsia will take to the new possibilities or, conversely, what sort of demands it will make on library services, and what sort of pressure the availability of new forms of publishing, funding or teaching will put on them. One thing we can say with confidence: unless there is a change in government policy towards a more consistent support, both financial and legislative, for education and research, these changes will appear slowly and take on many backhanded twists and turns.

At the beginning of this paper I argued that in discussing the introduction of new technologies, specifically information technologies, it is important to pay attention to the point of transition, to see all that is involved in this change of habit and why it is not just a "matter of time." The body of this paper, I hope, provided at least a glimpse of some of the friction points involved. For the time being, the last word, like the first, belongs to an economist, in this case to Vaclav Klaus, the prime minister of the Czech Republic, whose opinions expressed in a recent op-ed piece on "Science and our Economic Future," make him sound like a someone who has just bitten into a blue tomato only to find that it tastes like a peach.

> ...science is not about information, but about knowing, about thinking, about the ability to generalize thoughts, make models of them and then testable hypotheses that are to be tested. Science is not about the Internet and certainly not about its compulsory introduction. (Klaus, 1997)

NOTES:

[1] These and other features--such as WWW access to the catalogue--should be available shortly. As for the size of the database, of the total collection only a fragment is presently on-line. Prior

to the introduction of the new system, libraries had been cataloguing in ISIS. These records have been converted with little or no loss to the UNIMARC format which meant--in the case of the National Library--that, from the outset, several hundred thousand records were Aleph-ready. New acquisitions are catalogued directly into the new system and more records are made available through retrospective conversion (see below).

[2] For additional details on the this project, see the LINCA proposal presented to the Mellon foundation (LINCA, 1994).

[3] In fact, as of January 1997, the two campuses of UPJS have been redefined as two universities. The move was political (divide and conquer), playing off existing institutional rivalry. What the consequence of this will be on the project is not yet clear. The details of the original project can be found in the KOLIN proposal to the Mellon Foundation (KOLIN, 1995).

[4] For a more detailed account of the compatibility and conversion problem as well as of the solution, see Appendix H. of the MOLIN proposal (MOLIN, 1996). TinLib is the most widely used library system among Czech universities (the Czech vendor is located at Charles University in Prague).

[5] More precisely, out of the approximately 1.5 million volumes deposited in the central library (the Klementinum) about 1/5 were unshelved. Because these were new acquisitions--most in demand by users--most requests went unfulfilled.

[6] Hence also the symbolic significance of including the call number on the electronic record - it actually corresponds to a retrievable object! What a treat!

[7] There is no comparable library reconstruction going on in Slovakia, and it is not clear whether any of the authorized building projects in the Czech Republic will actually be funded, considering the most recent austerity measures introduced by the conservative government in the Spring of 1997.

[8] Equally important in the area of rare book and manuscript preservation is the direct digitalization project at the National Library in Prague, in which early medieval illuminated manuscripts are being scanned and made available on CD-ROM (in 1995/96: *Antiphonarium Sedlecense* and *Chronicon Concilii Constantiniensis*). This is a UNESCO sponsored *Memory of the World* project.

[9] The plan is to preserve the card catalogues in the library's archive and make use of the several rooms that they presently occupy as reading rooms.

[10] There is much more to this fascinating and complex project, well worth a separate study. The interested reader may wish to look at the original text of the project as it was presented to the Mellon Foundation for Funding (RETROCON, 1994), and at a special publication of the National Library devoted to this topic (Bares and Stoklasova, 1995).

[11] For the theoretically minded reader: I have put some of the terms in this paragraph in quotes on purpose, in order to highlight their own ambiguity in the anthropological scheme of things. The expression *cultural misunderstanding* is taken from a study with the same title and on that

very topic (Raymonde Carroll, 1987). I use *culture* as a cover term for what would be better served by the more obscure terms *habitus* or *discursive practice*. Most importantly, the use of the term is not meant to imply ethnic or national specificity or some bounded system of traditions. I borrowed the term *localization*, used to describe the adjustment that a library system software undergoes for a particular language, as a suitable characterization of the overall process in which one library tradition is made to fit a new (one could even say foreign) setting. Finally, *external ties* are often found embedded inside the organization. I wish to exercise caution in using this term since it is often quite difficult to pinpoint where an organization ends and the external world begins.

[12] The role of time management and, in particular, of delays in the implementation of the library project is a topic of a separate study (Lass, n.d.).

[13] According to a recent document (issued by the Slovak Ministry of Culture), the library's new mandate would include, among other things, the issuing of ISBN and the development of the national bibliographic records. This has put them in the situation, apparently desired, of having to demand the transfer of positions, computer hardware (mostly CASLIN Mellon purchases), and existing databases from the Slovak National Library--it remains unclear how they would gain the expertise--without which they cannot do the job. While there are other examples in which institutional rivalries have adversely affected the CASLIN project, in only some of them does the rivalry reside with the libraries themselves. In several instances it is the libraries that are caught in the middle of a battle. Such is the case in the KOLIN project discussed above (see also footnote 3.).

[14] The symbolic significance of foreign (Western) funds is not to be underestimated, nor should the role that this phenomenon has on the actual implementation of the project (covered best by the anthropological studies of 'cargo cults' or witchcraft). As for the politics of institutional positioning, the situation under review was made transparent (and more complicated) by the break-up of Czechoslovakia and, following that, by the surfacing of other regional tensions. As a result, the relationship across the border is more amicable (there is nothing to compete over), while the relationship between the two libraries in each of the countries is much less so.

[15] One of the surprises was the funding of automation at the National (then University or State) Library during the 1980's in Prague, which resulted in the development of a local machine-readable record format (MAKS) that became the accepted system among a majority of Czech and Slovak libraries. The grounds for "technology transfer" were therefore prepared (contrary to some who maintained that there was no expertise in place) when automation arrived in earnest after 1990.

[16] Ironically, the library became one of the "safe places" to "hide" politically discredited intellectuals (from the post 1968 purges).

[17] If the purchase of foreign (especially Western) books and periodicals was restricted for mostly political reasons, it is now actually stopped altogether due to zero (!) funding.

[18] As a result, organizational behavior retains its characteristic sluggishness. It is reasonable to predict that as the constraints on the budget continue to increase, so will the familiar ability to "trick the system." Under these conditions, teaching new management skills has been close to impossible and introducing a new record structure and cataloguing rules very slow. This

accounts, to a large extent, for the continued backlog of uncatalogued books.

[19] The fact that 'planning' is a discredited term doesn't help. And trying to explain that socialist planning and strategic planning may be quite different things doesn't seem to work.

[20] On the Czech side, the Ministry of Culture has decided to support library automation projects throughout the country, in the form of capital investment grants (no funds for salaries), that meet CASLIN standards.

REFERENCES:

Bares, Miroslav and Stoklasova, Bohdana

1995 *Retrospective Conversion in Czech Libraries*, Praha: Narodni Knihovna

Carroll, Raymonde

1988 *Cultural Misunderstandings: The French-American Experience*, Transl. Carol Volk, Chicago: Chicago University Press

KOLIN

1995 *Kosice Library Information Network*, A proposal presented to the Andrew W. Mellon Foundation, South Hadley, MA.

Klaus, Vaclav "Veda a nase hospodarska budoucnost." In: *Lidove noviny*, May 31, 1997.

Lass, Andrew

n.d. "Managing Delay: The Micropolitics of Time in the Czech and Slovak Library Automation Projects." Paper to be presented at Mellon Foundation Conference on Library Automation, Warsaw, October 16-18, 1997. [in progress]

LINCA

1994 *Library Information Network of the Czech Academy of Science*, A proposal presented to the Andrew W. Mellon Foundation, South Hadley, MA.

MOLIN

1996 *Moravian Library Information Network*, A proposal presented to the Andrew W. Mellon Foundation, South Hadley, MA.

RETROCON

1994 *National Library in Prague: retrospective catalogue conversion*, A proposal presented to the Andrew W. Mellon Foundation, South Hadley, MA.

Quandt, Richard

1996 "Electronic Publishing and Virtual Libraries: Issues and an Agenda for The Andrew W.Mellon Foundation," *Serial Review*, Summer 1996, pp.9-24.

……

For additional information about the conference, or The Andrew W. Mellon Foundation's scholarly communication initiatives, please contact Richard Ekman. For additional information about ARL or this web site contact Patricia Brennan, ARL Program Officer at (202) 296-2296.

Return to Office of Scholarly Communication Home Page

ARL Home ARL Scholarly Communication and Technology Home Page

© Association of Research Libraries, Washington, DC

Web Design by Angelo F. Cruz
Maintained by ARL Web Administrator
Last Modified: August 29, 1997

Scholarly Communication and Technology

Conference Organized by The Andrew W. Mellon Foundation

at Emory University
April 24-25, 1997

Copyright © of the papers on this site are held by the individual authors or The Andrew W. Mellon Foundation. Permission is granted to reproduce and distribute copies of these works for nonprofit educational or library purposes, provided that the author, source, and copyright notice are included on each copy. For commercial use, please contact Richard Ekman at the The Andrew W. Mellon Foundation.

Session #7 Multi-Institutional Cooperation

Consortial Access Versus Ownership

Richard W. Meyer
Director of Libraries
Elizabeth Coates Maddux Library
Trinity University

rmeyer@trinity.edu

……

The author wishes to acknowledge with thanks the financial support of The Andrew W. Mellon Foundation and the participation of several colleagues from libraries of the Associated Colleges of the South. Thanks also to my associate Tanya Pinedo for data gathering and analysis. All errors remain the property of the author.

Introduction

This paper reports on a consortial attempt to overcome the high costs of scholarly journals and to study the roots of the cost problem. The advent of high-speed telecommunication networks linking scholarly research throughout the world offers opportunity for reducing the costs to libraries for scholarly communications. The literature on the problem of journal costs includes both proposals for new ways of communicating research results as well as many studies on journal pricing.

Prominent members of the library profession have written proposals on how to disengage from print publishers.[1,2] Others in the sciences have suggested that electronic publications soon will emerge and bring an end to print-based scholarship.[3,4] Another scientist proposes that libraries solve the problem by publishing journals themselves.[5] These proposals, however, tend not to accommodate the argument that loosely coupled systems cannot be easily restructured.[6] While access rather than ownership promises cost savings to libraries, the inflation problem requires further analysis of the factors that establish journal prices before it is solved.

Many efforts to explain the problem of high inflation occupy the literature of the library profession and other disciplines. The most exhaustive description of the problem to date, published by the Association of Research Libraries for the Andrew W. Mellon foundation, provides ample data, but no solution. [7] Examples of the problem appear frequently in the *Newsletter on Serials Pricing Issues*, which was developed expressly to focus discussion of the issue.[8] Searches for answers appear to have seriously started with the studies of Hamaker and Astle, who provided a partial explanation of the problem based on currency exchange rates that work against libraries in North America.[9,10,11] Analyses published by librarians and economists propose means to escape inflation, which include securing federal subsidies, complaining to publishers, raising photocopying charges and convincing institutional administrators to increase budgets.[12,13,14,15]

A significant number of pricing analyses in recent years attempt to isolate the variables which determine prices and the difference in prices between libraries and individuals. Studies typically examine price by looking at the statistical relevance of sundry variables, but especially publisher type.[16,17,18] They confirm the belief of librarians that certain publishers, notably in Western Europe, practice price discrimination.[19,20,21] They also show that periodical prices are driven by many factors, including cost of production, which is related to frequency of issue, number of pages, and presence of illustrations. Alternative revenue from advertising and exchange rate risk for foreign publishers also affect price.[22,23] Quality measures on the content, such as number of times a periodical is cited affects demand, which then impacts price. Economies of scale that are available to some journals with large circulation affects price also.[24] These articles also help explain price differentials between what individuals are charged versus what libraries pay.[25] Revenues lost to photocopying accounts for much of the difference between individual and library price.[26] Also, differences in the way electronic journals may be produced compared to print provides a point on which some cost savings could be based.

The costs of production and the speed of communication may be driving forces that determine whether or not new publications emerge in the electronic domain to replace print. However, this issue involves a broader set of considerations. In a framework shaped by government policy, the interaction of demand and supply, more than the costs of production or speed of delivery, determines the price of any given journal. Periodical prices remain quite low over time when magazine publishers sell advertising as the principal generator of revenue, because publishers compete for readers, whose numbers can be sold to advertisers, rather than for the reader's dollars. When for political, public relations or similar reasons, publication costs are borne by organizations, usually other than scholarly societies, periodical prices tend to be lower. Prices tend to inflate in markets with high demand, where publishers are involved in supporting the communication of scholarly output. The highest demands and prices are concentrated in the sciences where multiple users include practicing physicians, pharmaceutical firms, national laboratories and so forth. Scholarly publishing in the sciences where demand is high provides the focus for much of the study of pricing and efforts to control library costs.

Unfortunately for libraries, the demand from users for any given scholarly journal is usually inelastic. Libraries tend to retain subscriptions no matter how high the publisher raises the price, because the demand originates with non-paying users even though libraries pay the bills. In turn, user demands are driven by price increases charged to individual subscribers to scholarly journals. Therefore, it might be expected that as currently existing print publications are offered by publishers in an electronic form, they will retain both their price as well as inelastic demand. Commercial publishers, who are profit maximizers, will seek to retain or improve their profits when expanding into the electronic market. However, there are some properties associated with electronic journals that could relax the inelasticity of journal prices. Diminished need for the physical artifact character of journals combined with changes in the transactions process to acquire scholarly content in the electronic domain could offset the profit potential of traditional scholarly publishing.

This paper reports on a multi-discipline study of the impact of electronic publishing on the pricing of scholarly periodicals. A brief overview of the pricing issue comparing print and electronic publishing is followed by a summary of the access approach to cost containment technique. This is then followed by a preliminary report on an attempt at this technique by a consortium and on the associated econometric study.[27]

Overview of Pricing Relevant to Electronic Journals

The industry of scholarly print publishing falls into the category of monopolistic competition, which is characterized by the presence of many firms with differentiated products, and by no barriers to entry of new firms.[28,29] Commercial and societal publishers supply a set of heterogeneous products which are distinguished from each other by quality and by uniqueness of content. Variation in quality occurs not only within any given journal, since articles differ somewhat in quality, but also from title to title. Furthermore, each scholarly article is fundamentally unique and has no perfect substitutes. As a result of this product differentiation, scholarly publishers do not encounter perfectly elastic aggregate demand typically associated with competitive markets. Rather, each publisher perceives a negatively sloped individual demand curve. Therefore, at least partially, each supplier has the opportunity to control the price of its product, even though barriers to entry of new, competing periodical titles may be quite low. Given this opportunity, publishers have gradually raised their prices to libraries with some loss of sales, but with consequent increases in profits which overwhelm those losses. They segment their market between individuals and libraries and charge higher prices to the latter in an effort to extract consumer surplus.

As publishers have lost sales of periodicals to individuals, scholars have increased their dependency on libraries, which in turn, have increased interlibrary borrowing to secure the articles needed by their users. The photocopies typically supplied via library collections represent some of the revenue potentially lost to publishers, but which is recaptured in the price differential. Although copyright protection and diligence of librarians replaces some lost revenue through copyright clearance fees, additional revenue might be captured by publishers if they could effectively offer their products in online, electronic databases where they could monitor all duplication. This potential may rest on the ability of publishers to retain control in the electronic domain of the values they have traditionally added to scholarship.

Scholars demand of journals-- as in the economic sense of acquiring -- both input in the form of documentation of the latest and most accurate knowledge and/or information on scholarly subjects as well as outlets for their contributions to this pool of scholarship. They pay the following costs to deliver their output through print publishing: sometimes page charges; labor in creative and editorial effort; and usually, they relinquish copyright in trade for acceptance of their scholarly efforts.

In exchange for their trade of copyright, scholars receive value in four areas. First, scholars secure value in communication, when every individual's contribution to knowledge is conveyed to others; thus impacting the reputation of the author's future output and educating the reader, which is input to the scholar's peers. Second, although not provided by publishers directly, archiving (traditionally, storage of print publication) provides value by preserving historically relevant scholarship and fixing it in time. This value arises essentially automatically as a consequence of storing physical artifacts in libraries. Third, great value accrues from filtering of contributions in given disciplines by separating them into levels of quality, which improves search costs allocation and establishes or enhances reputation. Fourth, segmenting of scholarship into discipline groupings is important in reducing input search costs to scholars, but at some expense to publishers who bear production costs. This exchange of copyright ownership for value could be dramatically affected with the emergence of electronic journals.

Electronic journals are emerging in two ways. Totally new titles are appearing exclusively in electronic form in order to take advantage of the speed and informality of the electronic environment. Alternatively, existing print titles are being transformed or augmented by electronic counterparts as publishers look to improve marketability. Some new journals have begun exclusively as electronic publications with mixed success. The directory published by the Association of Research Libraries listed approximately 27 new electronic journals in 1991. By 1995 that figure had risen to over 300, of which some 200 claim to be peer reviewed.[30] Since then hundreds more electronic journals have been added, but the bulk of these additions appear to be electronic counterparts of previously existing print journals.[31] Constraints may keep many of these from succeeding.

The infrastructure and inter-relationships of scholarly print publishing evolved over a long time. In order for a parallel structure to emerge in the electronic domain, electronic publishers have to add as much value to the process of scholarship as they do in print. Value must be added in archiving, filtering and segmenting, in addition to, communication. It is essential that electronic products establish a brand name that readily communicates their level of quality. Traditionally, the reputation of editors establishes brand name which rests on and must be nurtured by years of consistent performance. While some new scholarly titles are emerging successfully, traditional publishing retains an edge in the electronic domain.

Two of the more successful electronic journals of interest to librarians have not performed as well as hoped. *PACS Review*, which is a widely distributed publication from the University of Houston on electronic catalogs, shows a trend in new submissions per year that is flat at best and more likely declining. Over the five year period 1990 to 1995, the number of articles in *PACS Review* declined from 16 to 5; the number of pages from 241 to 78. As well, the number of new authors declined. Further examination of the titles cited in the publication suggests a drop in interest, also. The first volume contained original articles on a variety of topics. By the third and fourth volumes, submissions were more like reprises of conference papers. In 1996, interest may have rebounded somewhat with several substantial contributions.

As another interesting example, the electronic publication called *EJournal*, proclaims itself an "electronic journal concerned with the implications of electronic networks and texts" but showed erratic publication output, from a high of several thousand lines and five articles in its second year to a low of one article with less than one thousand lines in the fifth year. This publication appears to also suffer from submission problems, especially since more than one issue has solicited articles from readers.[32] A similar story could be written for many of the other electronic attempts. Empirical work indicates that electronic publications are inconsequential to date and that no more than three electronic journals have had substantive impact on scholarship.[33]

The apparently mixed success of new titles derives from the endemic need to provide the values traditionally added by publishers. Establishing brand quality requires tremendous energy and commitment. There are some successful electronic titles sponsored by individuals who are fervent in their efforts to demonstrate that the scholarly community can control the process of communicating scholarship. However, it is obviously unrealistic to expect an instantaneous, successful emergence of a full-blown infrastructure in the electronic domain that overcomes the obstacles to providing the values required by scholars. The advantage of higher communication speed of electronics is insufficient to drive a transformation of scholarly communication quickly.

In contrast, it appears likely that a transformation from print to electronic publication will be achieved effectively by duplicating existing print journals in the electronic sphere. Publishers of established print journals face less imposing investments to add electronic counterparts to their product lines. Traditional print journals are being packaged into collections and successfully marketed to libraries in electronic form. For example, the *Adonis* [34] collection on CD-ROM contains over 600 long-standing journals in medicine, biology and related areas covering about seven years. Furthermore, Ebsco, University Microfilms (UMI), Information Access Company (IAC), Johns Hopkins University Press, OCLC and other companies are implementing similar products. OCLC now offers libraries access to the full-text of journal collections pulled together by UMI and Ebsco. Furthermore, Johns Hopkins is making all forty plus titles which that press publishes now available online through *Project MUSE*.

Publications already existing in print are at least two steps ahead of any new electronic title on the pathway to complete transformation. Costs and values associated with filtering, segmenting and archiving that must be considered in addition to communicating, appear to be overcome by existing journals that are migrating to electronic form.

During the past fifteen years, libraries have experienced a remarkable shift from acquiring secondary sources in print to accessing them through a variety of electronic venues. Users of most academic libraries today find CD-ROM indexes, local online indexes and electronic gateways over the Internet to indexes on remote servers. Many librarians report that patrons seldom use print indexes any more. In effect, much of the secondary literature has already made the transformation from print to electronic. In this environment, cost per unit of information delivered has often declined dramatically, because user costs of seeking information in the form of labor have diminished, thereby raising the use rate of indexes.[35] Presumably, these efforts were cost effective because they reduced the time needed by library users locating information and because they have proven to be more powerful retrieval agents due to Boolean logic and diminished need for thesaurus control.

This phenomenon suggests that many scholarly periodicals will become available electronically as an automatic response to the economies available there. In fact, there are quite a few products emerging which offer electronic bundles of periodical titles on given disciplines or general interest. Some of these represent viable possibilities for shared access among a consortium of libraries, with consequent savings from cancellation of print subscriptions.

Pricing of Electronic Journals

Some monopoly power of publishers could be lost if barriers to the entry of new journals are lower in the electronic domain than in the print domain. With full-text online, libraries may take advantage of the economies of sharing access, which electronic networks offer. Favorable economies come into play when a group of libraries contracts for shared access to a core collection. Sharing a given number of access ports allows economies of scale to take effect. Were one access port each provided to a consortium of fifteen libraries, the vendor would tie up a total of fifteen ports, but any given library in the group would have difficulty servicing a user population with one port. Whereas for example, by combining access, fifteen libraries together might get by with as few as ten ports collectively. The statistical likelihood is small that all ten ports would be needed collectively by the consortium at any single given moment. This saves the vendor some computer resources that can then lead to a discount for the consortium that nets out less cost to the libraries. For example, fifty members of the Oberlin Group of college libraries negotiated a contract for all the periodicals of the Johns Hopkins Press *Project Muse* for a fifty-percent discount from their electronic list price.

Although numerous models for marketing exist, such as bundling CD-ROM's into the subscription or giving discounts for advanced deposits toward article purchases, there are fundamentally only two ways that publishers can price their products in the electronic domain. Either they will offer their products on *subscription* to each title or group of titles for a flat fee, or they will price the content on an article-by-article *transaction* basis. Vendor collections of journals for one flat fee based on the size of the user population represents a variant on the subscription fee approach. Commercial publishers, who are profit maximizers, will choose the method with the higher potential to increase their profit. Transaction based pricing offers the possibility of capturing revenue lost to interlibrary lending. Also, demand for content could increase due to the ease of access afforded online. On the risk side, print subscription losses would occur where the cumulative expenditure for transactions from a given title is less than its subscription price.

One or both of two mechanisms potentially could flatten demand functions in the electronic domain. First, by making articles available individually to consumers, the separation of items of specific interest to given scholars creates quality competition that increases the elasticity of demand, because quality varies from article to article. Presumably, like individual grocery items, the elasticity of demand for particular articles is more elastic than that of periodical titles. A trip to the grocery store involves buying groceries in general and buying specific groceries. Economists argue that the demand for tortillas is more elastic than for groceries in general because other bakery goods can be substituted. Whereas, there is no substitute for groceries in general except higher priced restaurant eating. Similarly, it may be argued that when faced with buying individual articles, price increases will dampen demand more quickly than would be the case for a bundle of articles which are of interest to a group of consumers.

Second, by offering articles in an environment where the consuming scholar is required to pay directly (or at least observe the cost to the library), the effect of separation of payer and demander common with library collections resulting in high inelasticity will be diminished. Combining payer and consumer will increase elasticity because scholars will no longer be faced with a zero price. Even if for some libraries the scholar is not constrained to pay directly for the article, increased awareness of price will have a dampening effect on inelasticity. However, publishers may find it possible to price individual articles at a level that cumulatively exceeds the price they are able to set for a journal title which bundles a group of articles together. That is, the sum of individual article fees paid by consumers may exceed the bundled subscription price formerly experienced by libraries forced to purchase a whole title to get individual articles in the print realm.

For a product like *Adonis*, which is a sizeable collection of periodicals in the narrow area of biomedicine, transaction based pricing works out in favor of the consumer versus the provider. This is because there will likely be only a small number of articles of interest to consumers from each periodical title. This makes purchasing one article at a time more attractive than buying a subscription, because less total expenditure will normally result. In the case of a product composed of a cross section of general purpose periodicals such as the UMI *Periodical Abstracts* full-text product, the opposite may be true. The probability is higher that a user population at a college may collectively be interested in every single article in general purpose journals. This makes subscription based pricing more favorable for libraries, because the cumulative-cost of numerous transactions could easily exceed the subscription price. Publishers will seek to offer journals in accordance with whichever of these two scenarios results in the higher profit. Scientific publishers will tend to bundle their articles together and make products available as subscriptions to either individual journals or groups. Scholarly publishers with titles of general interest will be drawn toward article by article marketing.

An Elsevier effort to make 1,100 scientific titles available electronically will be priced on a title by title subscription basis and at prices higher than the print version when only the electronic version is purchased.[36] On the other hand, the general purpose titles included in UMI's *Periodical Abstracts* full-text, (as are the similar products of Ebsco and IAC), as an alternative interface to their periodicals, are available on a transaction basis by article. These two approaches seek to maximize profit in accordance with the nature of the products.

Currently, UMI, Ebsco, and IAC, who function as the aggregators, have negotiated arrangements that allow site licenses for unlimited purchasing. These companies are operating as vendors who make collections of general purpose titles available under arrangements that pay the publishers royalties for each copy of their articles printed by library users. UMI, IAC and Ebsco have established license arrangements with libraries for unlimited printing with license fees based on expected printing activity. These arrangements offer some libraries a solution to the fundamental pricing problem of monopoly power by publishers.

New research could test whether publishers are able to retain monopoly power with electronic counterparts to their journals. Work using an alternative model has examined the possibility that publishers exercise monopoly power in setting prices. Theory predicts that in a competitive market, even when it is characterized as monopolistic competition, the price offered to individuals will tend to remain elastic. Faced with a change in price of the subscriptions purchased from his own pocket, a scholar will act discriminately. Raise the price to individuals and some will cancel their subscriptions in favor of access to a library. That is, the price of periodicals to individuals is a determinant of demand for library access. By substituting a measure of monopoly power in place of price, it has been shown that publishers have some ability to influence their earnings through price discrimination.[37]

In contrast, the price to libraries, which is often much higher than the price to individuals, is set at a level intended to extract consumer surplus. The difference in these prices provides a reasonable measure of the extent of that monopoly power, assuming that the individual subscription price is an acceptable proxy for the marginal cost of production.[38] Even if not perfect, some measure of monopoly power is represented by the difference in prices. Extending this line of research may show that monopoly power is independent of the medium.

In monopolistic competition, anything which differentiates a product may increase monopoly power. Firms that sell laundry detergent expend tremendous amounts of money on advertising. They do so to create the impression that their product is qualitatively distinguishable from others. It may be that electronic availability of specific titles will create an impression of superior quality that could lead to higher prices. However, the prices of journals across disciplines also may be driven by different factors. In general, prices are higher in the sciences and technical areas and lower in the humanities. This is understandable considering the market for science versus humanities. There is essentially no market for scholarly publications in the humanities outside of academe, whereas scientific publications are used heavily in corporate research by pharmaceutical firms and other industries highly dependent on research. As a result, monopoly power will likely be demonstrable in the sciences, but not in other general areas. This would reflect additional price discrimination in the electronic environment by publishers who are able to capture revenue lost to photocopying.

Access Versus Ownership Strategy

Clearly, if commercial publishers continue to retain or enhance their monopoly power with electronic counterparts of their journals, the academic marketplace must adjust or react more effectively than it has in the past. Possibly, the reaction of universities could lead to erosion of previous success achieved with price discrimination if an appropriate strategy is followed. Instead of owning the periodicals needed by their patrons, some libraries have experimented with replacing subscriptions with document delivery services. Louisiana State University reports cancelling a major portion of their print journals.[39] They replaced these cancellations by offering faculty and students unlimited subsidized use of a document delivery service. The first year cost for all the articles delivered through this service was much less than the total cost to the library for the former subscriptions. Major savings for the library budget via this approach would appeal to library directors and university administrators as a fruitful solution. However, it will turn out to be short term at best.

Carried to its logical conclusion, this approach produces a world in which each journal is reduced to one subscription shared by all libraries. This is equivalent to every existing journal having migrated to single copies in online files accessible to all interested libraries. Some libraries will pay a license fee in advance to allow users unlimited printing access to the online title and some libraries will require users to pay for each article individually. This requires the entire fixed-cost-plus-profit components of publisher's revenue to be distributed over article prints only. Whereas, with print publications, the purchase of subscriptions of physical artifacts that included many articles not needed immediately, brought with it a bonus. The library acquired and retained many articles with future potential use. Transactions based purchasing sacrifices this bonus and increases the marginal cost of articles in the long run.

Put another way, the marginal cost of a journal article in the print domain was suppressed by the spread of expenditure over many items never read. In the electronic domain under transactions based pricing, users face a higher, more direct price and therefore are more likely to forego access. While the marginal benefit to the user may be equivalent, the higher marginal cost makes it less likely users will ask for any given article. The result may show up in diminished scholarly output or notably higher prices per article.

More likely in the long-term, should a majority of libraries take this approach, it carries a benefit for publishers. There has been no means available in the past for publishers to count the actual number of photocopies made in libraries and thus to set their price accordingly. The electronic domain could make all those hidden transactions readily apparent. As a result, publishers could effectively maintain their corporate control of prices, and do so with more accurate information with which to calculate license fees. Given this attempted solution, publishers would be able to regain and strengthen their monopoly position.

A more promising approach lies in consortial projects such as that conducted by the Associated Colleges of the South (ACS).[40] There are collections in full-text of over 1,000 existing journals with backfiles accompanying the *Periodical Abstracts* and *ABI/Inform* indexes of UMI. These are available directly online from the vendor or through OCLC. The ACS contracted an annual license for these two products for the thirteen schools represented. Similar to the cost for each ACS library, the cost to Trinity University is $11,000 per year in 1996-97 for the electronic periodicals in the UMI databases. Coincidently, Trinity University subscribes to the print version of 373 titles covered by these products. Trinity could cancel its subscriptions to the print counterparts of the journals provided, and save $24,900. Although Trinity's library will subsidize user printing for paper, toner, and so forth, at an expected cost of several thousand dollars per year, with 230 faculty and 2,400 students, it appears likely that favorable economies accrue from switching to these electronic products. Of course, these savings will be accompanied by a significant decrease in non-dollar user cost to patrons, so unmet demand will emerge to offset some of the savings. Moreover, there is a substantial bonus for Trinity users inherent in this arrangement.

There is a number of titles made available in the UMI product for which subscriptions would be desirable at Trinity, but which have not been purchased in the past, because of budget limitations. There are some of these from which users would have acquired articles through the normal channels of interlibrary loan. However, the interlibrary loan process imposes costs in the form of staff time and also user labor and is sufficiently cumbersome that many users avoid it for marginally relevant articles. However, if some of those marginal articles could be easily viewed on screen as a result of electronic access described in this example, some users would consider the labor cost of acquiring them to have been sufficiently reduced to encourage printing the articles from the system. Therefore, the net number of article copies delivered to users will be significantly increased simultaneous with a substantial net decrease in the cost of subscriptions delivered to libraries.

Included in this equation are savings which accrue to the consortial libraries by sharing access to electronic subscriptions. Shared access will result in a specific number of print cancellations which will decrease publisher profit from subscriptions. Publishers offering their journals in the electronic domain will be confronted by a change in the economic infrastructure that will flatten the scholar's demand functions for their titles while simultaneously increasing the availability of articles to the direct consumers. By lowering the user's non-dollar cost of accessing individual articles, demand will increase for those items. Scholars, therefore, will be more likely to print an article from an electronic library than they would be to request it through interlibrary loan. However, depending on library policy, those scholars may be confronted with a pay per print fee, which will affect their demand function. If the publisher raises the price to scholars for an article, they are more liable to lose a sale. Users will be more cautious with their own money than with a library's. This is to say that in the electronic domain, where scholars may be paying directly for their consumption, demand functions will be more elastic. This will occur to some extent even when users do not pay for articles, but merely note the article price paid by their subsidizing library. Therefore price discrimination may be more difficult to apply and monopoly power will be temporarily lost.

The loss might be temporary, because this strategy is functionally the same as merging several libraries into one large library and providing transactions based access versus ownership. This super library could ultimately face similar price discrimination currently existing in the print domain. This will lead, in turn, to the same kind of inflation that has been suffered for many years.

Preliminary Analysis of Financial Impact

This paper reports on the early stages of a three-year study funded by the Andrew W. Mellon Foundation. This study is collecting data on approximately 6,000 journal titles gathered from the combined subscription lists of the thirteen ACS libraries. The study includes analysis directed at testing the viability of consortial access versus ownership as well as the potential long term solution that would derive from emergence of a new core of electronic titles. A complete financial analysis of the impact of consortial, electronic access to a core collection of general purpose periodicals as well as an econometric analysis of the impact of electronic availability on pricing policy will issue from the study conducted under this grant. Some interesting issues have emerged with preliminary results of the study.

Financial Analysis

The Palladian Alliance is a project of the Associated Colleges of the South funded by the Andrew W. Mellon Foundation. This consortium of thirteen liberal arts colleges -- not just libraries -- has a full time staff and organizational structure. The Palladian Alliance came about as result of discussions among the library directors who were concerned about the problem described in this paper. As the project emerged, it combined the goals of several entities, which are shown in Table 1 along with the specific objectives of the project.

Andrew W. Mellon Foundation awarded a grant of $ 1.2 million in December 1995 to the ACS. During the first half of 1996, the librarians upgraded hardware, selected a vendor to provide a core collection of electronic full-text titles, and conducted appropriate training sessions. Public and Ariel workstations were installed in libraries by July 1996 and necessary improvements were made to the campus networks to provide access for using world-wide web technology. Training workshops were developed under contract with Amigos and SOLINET on technical aspects and were conducted in May 1996. During that same time, an analysis was conducted to isolate an appropriate full-text vendor.

After comparison of the merged print subscription list of all institutions with three products -- IAC's InfoTrac, Ebsco's EbscoHOST, and UMI's *Periodical Abstracts* and *ABI/Inform* -- the project team selected UMI with access thru OCLC. A contract with OCLC was signed in June for July 1, 1996 start-up of FirstSearch for the nine core databases: *WorldCat, FastDoc, ERIC, Medline, GPO Catalog, ArticleFirst, PapersFirst, ContentsFirst, ProceedingsFirst*; and for University Microfilm's two core indexes: *Periodical Abstracts* and *ABI/Inform* along with their associated full-text databases. This arrangement for the UMI products provides a general core collection with indexing for 2,600 titles of which approximately 910 also provide full-text of the contents.

Table 1. Goals and Objectives of the ACS Consortial Access Project.

Goals of the ACS Libraries:

- Improve the quality of access to current information
- Make the most efficient use of resources

Goals of the ACS Deans:

- Cost Containment

Goals of the Andrew W. Mellon Foundation:

- Relieve the economic pressure from periodical price inflation
- Evaluate the impact of electronic access on publisher pricing practices

Objectives of the Project:

- Improve the hardware available within the libraries for electronic access
- Provide online access to important undergraduate periodical indexes
- Provide online access to core undergraduate periodicals in full text
- Provide campus-wide access through readily available search tools -- eg., Internet browsers such as Netscape
- Determine the financial impact on the ACS libraries
- Test the pricing practices of publishers and their monopoly power

The UMI via OCLC FirstSearch subscription was chosen because it offered several advantages including a reliable, proprietary backup to the Internet, additional valuable databases at little cost, and easy means to add other databases. The UMI databases offered the best combination of cost and match with existing holdings. Most of the libraries had none of these databases. A few had UMI, Ebscohost or InfoTrac products.

Students have had access to the core electronic titles since the Fall semester in 1996. As experience builds, it is apparent that the libraries do have some opportunity to cancel print subscriptions with financial advantages. The potential costs, savings and added value are revealed in Tables 2 through 4. Specific financial impact on a few of the institutions during the first year are shown in Table 5. It should be noted that the financial impact is based on preliminary data that has been extremely difficult to gather. Publisher and vendor invoices vary considerably between schools on both descriptive information and prices. Therefore, these results will be updated continually throughout the project.

The following tables are based on actual financial information for the consortium. It should be understood that these figures do not include periodical titles acquired directly from publishers or gift subscriptions. Throughout these tables, it should be kept in mind that the data for Morehouse does not include the entire collection available at Atlanta University Center; this information will be updated later to give a more accurate description of the effect of the project at Atlanta. Table 2 summarizes the project costs. These calculations will be corrected to reflect revised enrollment figures immediately prior to renewal for the 2nd and 3rd years. The project was designed to use grant funds exclusively the first year, then gradually shift to full support on the library accounts by the fourth year.

Table 2. Cost Sharing Between the Grant and the Institutions.

Institution	Enrollment	% of Total Enrollment	First Year	Second Year	Third Year
Mellon Grant			$184,295	$120,705	$45,000
Atlanta	13,174	38.70%		$26,873	$61,917
Birmingham	1,406	4.13%		$2,868	$6,608
Centenary	821	2.41%		$1,675	$3,859
Centre	968	2.84%		$1,975	$4,550
Furman	2,673	7.85%		$5,452	$12,563
Hendrix	978	2.87%		$1,995	$4,597
Millsaps	1,278	3.75%		$2,607	$6,007
Richmond	3,820	11.22%		$7,792	$17,954
Rhodes	1,407	4.13%		$2,870	$6,613
Rollins	2,632	7.73%		$5,369	$12,370
Sewanee	1,257	3.69%		$2,564	$5,908
Southwestern	1,199	3.52%		$2,446	$5,635
Trinity	2,430	7.14%		$4,957	$11,421
TOTALS	34,043		$184,295	$190,147	$205,000

The ACS libraries collectively subscribe to approximately 14,200 subscriptions through their vendors as shown in Table 3. Of these, 6,000 are unique titles; the rest are duplicates of these unique titles. Were the ACS libraries collectively merged into one collection, it would therefore be possible to cancel over 8,000 duplications and save over $1,133,000. Since this is not possible, the libraries have contracted for electronic access to nearly 1,000 full-text titles from UMI, where over 600 UMI titles match the print subscriptions held by the collective libraries. Cancelling all but one of the print duplications of the UMI titles could save the libraries about $130,000 or cancelling all the print counterparts to the electronic versions would save about $185,000 which is approximately equal to the licensing costs for one year per Table 2.

Table 3. Potential Savings from Substitution of Online Full-Text for Print Subscriptions.

	No. Titles	Costs/Savings
Cost Total for All ACS Print Subscriptions	14,187	$2,017,565
Number of Unique Titles	6,073	$883,880
Number of Duplicate Titles	8,114	$1,133,685
Cancelling of All But One Overlapping Duplicates	2,269	$130,306
Cancelling of All Overlapping Duplicates	2,870	$185,395

The project adds considerable value to the institutional resources as a bonus. There are many titles available through UMI that the schools had not previously taken. Table 4 lists the number of print subscriptions carried by each institution and indicates how many of those are available in the UMI databases electronically. Were the print counterparts of all these electronic journals to be cancelled, the fourth column shows the savings available to each school. "Added E-titles" shows the number of new journals made available to each institution through the grant.

Table 4. Savings Potential for Each Institution and Value Added by Electronic Subscriptions.

Institution	No. Print Subs'ns	Overlap w/ UMI	Cancellation Savings	Added E-titles	Total Subs'ns
Birmingham	658	198	$13,583	712	1,370
Centenary	535	184	$10,831	726	1,261
Centre	790	194	$11,501	716	1,506
Furman	2,008	279	$17,632	631	2,639
Hendrix	573	180	$9,980	730	1,303
Millsaps	740	193	$12,425	717	1,457
Morehouse	49	41	$2,494	869	918
Rhodes	318	81	$4,248	829	1,147
Richmond	1,976	368	$25,315	542	2,518
Rollins	1,314	261	$19,078	649	1,963
Sewanee	1,607	214	$14,073	696	2,303
Southwestern	1,406	304	$19,333	606	2,012
Trinity	2,213	373	$24,903	537	2,750
TOTAL	14,187	2,870	$185,396	8,960	23,147

Table 5 details the financial impact on several ACS institutions. Comparing this table with Table 2 reveals that in the cases of Trinity, Millsaps, and Rollins, even without Mellon support, the consortial provision of the OCLC/UMI databases could be paid for by cancelling existing redundant indexes. In Trinity's case, two indexes previously purchased as CD-ROM's or direct links to another online source were cancelled for savings of over $5,000 in the first year. Trinity cancelled a CD-ROM subscription to a site license of *ABI/Inform* , which saved expenditures totaling over $6,000 per year and an online general purpose index that previously cost over $12,000. The Trinity share to the Palladian Alliance project would have been just over $13,000 per year for the first three years. Similarly, Millsaps cancelled one indexes and 74 periodical titles that overlapped the UMI content. Their first year savings were over $5,700.

Table 5. First Year Financial Impact on Selected ACS Schools.

* = Incomplete data, but no cancelations made.

	Birming'm	Centre	Hendrix	Millsaps	Rhodes *	Trinity
Periodical Subscriptions						
Total 1996	656	719	600	677	957	2686
Total 1997	660	723	604	633	512	2621
Cancellations						
Total for 1997	2	1	0	85	0	42
Overlap of UMI	0	0	0	74	0	0
Indexes	1	0	1	0	0	9
Savings						
Periodicals	$24	$120	$0	$9,274	$0	$20,049
Overlap of UMI	$0	$0	$0	$5,104	$0	$0
Print Indexes	$4,650	$0	$604	$0	$0	$7,806
Electronic Indexes	$0	$0	$0	$0	$0	$18,491
Savings Due to Project	$4,650	$0	$604	$5,104	$0	$26,297
Subsidized Cost of Project	$9,476	$6,524	$6,591	$8,613	$9,483	$16,378
NET SAVINGS	($4,826)	($6,524)	($5,987)	($3,509)	($9,483)	$9,919

The interesting outcomes of the project at this point include a couple of new pieces of important information. First, cancelling individual subscriptions to indexes provides a viable means to relieve campus budgets at least in a short run with consortial pricing. In Trinity's case, were it necessary to pay our full share of the cost, there were more than sufficient savings from cancelling indexes alone to pay for the project. The net savings over the project lifespan total nearly $18,000 for Trinity just considering trade-offs with indexes alone.

Second, on the down side, cancelling journals and replacing them with an aggragator's collection of electronic subscriptions may not be very reliable. It is apparent that the aggragators suffer from the vagaries of publishers. Over just the short time of the first few months of the project, UMI dropped and added a number of titles in both full-text databases. This means that instead of full runs of each title, there are often partial runs. Furthermore, in some cases, the publisher provides only significant articles, not the full journal. Therefore, the substitution of UMI provides the libraries with essentially a collection of articles, not a collection of electronic subscription substitutes. This diminishes reliability and discourages libraries from being able to secure really significant cost savings.

It should be noted however, that several of the libraries independently subscribed to the electronic access to Johns Hopkins *Project Muse.* In contrast to an aggregated collection, this project provides full-image access to every page of the print counterparts and guarantees access indefinitely to any year subscription once paid for. This means that reliability of the product is substantially improved and it provides reasonable incentives to the libraries to substitute access for collecting. While it may be acceptable to substitute access to a large file of general purpose articles for undergraduate students, *Project Muse* holds out better promise compared to the initial project for scholarly journal collections.

Third, the impact of online full-text content may or may not have an impact on interlibrary loan activity. Table 6 summarizes the searching and article delivery statistics for the first six months of the project compared to the total interlibrary borrowing as well as non-return photocopies ordered through the campus interlibrary loan offices. The change in interlibrary loan statistics for the first six months of the project compared to the previous year show that in some cases interlibrary borrowing increased and in other cases it decreased. Several variables in addition to the availability of full-text seem to affect use of interlibrary loan services. For instance, some of the institutions had full-text databases available before the project started. Some made more successful efforts to promote the project services than others. It seems likely that improved access to citations from online indexes made users more aware of items that could be borrowed. That effect probably offset an expected decrease in interlibrary loans that the availability of full-text makes predictable. Regardless, statistics on this issue yield inconclusive results early in the project.

Table 6. UMI Articles Delivered to Users Compared to Change in Interlibrary Loans from 1995 to 1996.

School	Enroll- ment	Total Searches	Searches per Student	Articles Delivered	Articles Delivered per Student
Atlanta	13,174		0.00		0.00
Birmingham	1,406	5,597	3.98	660	0.47
Centenary	821	789	0.96	0	0.00
Centre	968	5,541	5.72	1,583	1.64
Furman	2,673	8,065	3.02	373	0.14
Hendrix	978	1,392	1.42	498	0.51
Millsaps	1,278	11,175	8.74	4,523	3.54
Rhodes	3,820	1,744	0.46	384	0.10
Richmond	1,407	45,639	32.44	10,477	7.45
Rollins	2,632	8,862	3.37	3,052	1.16
Southwstrn	1,257	19,999	15.91	5,623	4.47
Sewanee	1,199	18,117	15.11	590	0.49
Trinity	2,430	66,218	27.25	25,835	10.63

	Non-returns 95	Non-returns 96	Change in Non Returns	Total Borrows 95	Total Borrows 96	Change in Total Borrowing
Atlanta						
Birmingham	662	668	0.91%	928	380	-59.05%
Centenary	583	441	-24.36%	911	1,137	24.81%
Centre	409	351	-14.18%	872	758	-13.07%
Furman	246	246	0.00%	833	923	10.80%
Hendrix	146	192	31.51%	251	353	40.64%
Millsaps	568	352	-38.03%	710	887	24.93%
Rhodes	255	198	-22.35%	601	471	-21.63%
Richmond	1,034	1,044	0.97%	1,892	1,831	-3.22%
Rollins	394	365	-7.36%	656	652	-0.61%
Southwstrn	412	308	-25.24%	695	571	-17.84%
Sewanee	626	434	-30.67%	1,083	1,038	-4.16%
Trinity	706	711	0.71%	1,172	1,257	7.25%

Econometric Analysis

At this point, a meaningful econometric analysis is many months away. It is intended that a model based on Lerner's definition of monopoly power will be used to examine pricing as journals shift into the electronic sphere. The model calls for regressing the price of individual titles on a variety of independent variables, such as number of pages, advertising content, circulation, publisher type, and including a dummy for whether a journal is available electronically or not. Data is being collected on over 2,000 of the subscriptions held by Trinity for the calendar years 1995- 1997. Difficulties with financial data coupled with the time-consuming nature of data gathering have delayed progress on the econometric analysis.

It would be desirable to conduct an analysis on time series data to observe the consequences in journal price changes as a shift is made to electronic products. This would provide a forecast of how publishers react. Lacking the opportunity at the outset to examine prices over time, a straightforward model applying OLS regression on cross section data similar to the analyses reported by others, will form the basis of the analysis. Earlier models have typically regressed price on a number of variables to distinguish the statistical relevance of publisher type in determining price. By modifying the earlier models this analysis seeks to determine whether monopoly power may be eroded in the electronic market. The methodology applied uses two specifications for an ordinary least squares regression model. The first, regresses price on the characteristics of a set of journal titles held by the ACS libraries. This dataset is considerably larger than those utilized in previous studies. Therefore, we propose to confirm the earlier works that concentrate on economic journals across a larger set of disciplines. This specification includes the variables established earlier: frequency of publication, circulation, pages per year, and several dummy variables to control for whether the journals contain advertising and to control for country of publication. Four dummy variables are included for type of publisher with the residual being commercial. A second specification regressing the difference in price for libraries compared to individuals will be regressed on the same set of variables with an additional dummy added to show whether given journals are available electronically or not.[41]

The ACS libraries collectively subscribe to approximately 14,000 titles. Where they duplicate, an electronic set has been substituted for shared access. We anticipate that at the margin, the impact on publishers of ACS cancelling subscriptions to the print counterparts of this set would be minimal. However, the national availability of the electronic versions will precipitate cancellations among many institutions in favor of electronic access. Prices will be adjusted accordingly. Since most publishers will offer some products in print only and others within the described electronic set, we expect the prices of the electronic version will reflect an erosion of monopoly power. Thus the cross section data will capture the affect of electronic availability on monopoly power.

Since the dataset is comprised of several thousand periodical titles representing general and more popular items, several concerns experienced by other investigators will be mitigated. The only study found in the literature so far that examines publishers from the standpoint of the exercise of monopoly power investigated price discrimination.[42] This project intends to extend that analysis in two ways. First, we will use a much broader database. Most of the previous work has been done on limited datasets of less than 100 titles narrowly focused in a single academic discipline. Second, we will extend the analysis by assuming the existence of price discrimination given the difference in price to individuals versus libraries for most scholarly journals. With controls in the model for previous discoveries regarding price discrimination, we will attempt to test the null hypothesis that monopoly power will not decrease in the electronic domain.

In the dataset available, we were unable to distinguish the specific price of each journal for the electronic replacement, because UMI priced the entire set for a flat fee. This pricing scheme may reflect an attempt by publishers to capture revenue lost to interlibrary lending. However, it may also reflect publisher expectations that article demand will increase when user non-dollar costs decrease. Therefore, monopoly power will be reflected back on to the subscription price of print versions. As a result we will use the price of print copies as a proxy for the specific electronic price of each title.

An alternative result could emerge. In monopolistic competition, anything which differentiates a product may increase its monopoly power. For example, firms that sell laundry detergent expend tremendous amounts of money on advertising to create the impression that their product is qualitatively distinguishable from others. It may be that electronic availability of specific titles will create an impression of superior quality.

The general model of the first specification is written:

$$\text{yj} = \alpha + \beta 1\text{IPRICEj} + \beta 2\text{CIRCj} + \beta 3\text{FREQj} + \beta 4\text{PAGESj} + \beta 5\text{AGEj} + \beta 6\text{QUALITYj} + \beta 7\text{PEERREVj} + \beta 8\text{CCCREGj} + \beta 9\text{ADVj} + \beta 10\text{ASSOCj} + \beta 11\text{GOVERNj} + \beta 12\text{FOUNDTNj} + \beta 13\text{UNIVPRj} + \beta 14\text{EUROPEj} + \beta 15\text{GBRITAINj} + \beta 16\text{OTHERj} + \beta 17\text{ELECTRNj} + \varepsilon\text{j}$$

where, y equals the library price (LPRICE) for journal $j = 1, 2, 3, \ldots$ n. The definitions of independent variables appear in Table 6 along with the expected signs on and calculations of the parameters $\beta 1$ through $\beta 17$ to be estimated by traditional single regression techniques.

The general model of the second specification is written:

$$\text{yij} = \alpha\text{i} + \beta 1\text{iRISKj} + \beta 2\text{iCIRCj} + \beta 3\text{iFREQj} + \beta 4\text{iPAGESj} + \beta 5\text{iAGEj} + \beta 6\text{iQUALITYj} + \beta 7\text{iPEERREVj} + \beta 8\text{iCCCREGj} + \beta 9\text{iADVj} + \beta 10\text{iASSOCj} + \beta 11\text{iGOVERNj} + \beta 12\text{iFOUNDTNj} + \beta 13\text{iUNIVPRj} + \beta 14\text{iEUROPEj} + \beta 15\text{iGBRITAINj} + \beta 16\text{iOTHERj} + \beta 17\text{iELECTRNj} + \varepsilon\text{ij}\ (1\text{-i})$$

where, y equals two different forms of monopoly power (MPOWER1; MPOWER2) defined as measure $i = 1$ and 2 for journal $j = 1, 2, 3, \ldots$n. The definitions of independent variables appear in Table 6 along with the expected signs on and calculations of the parameters $\beta 1$ through $\beta 17$ to be estimated by traditional single regression techniques.

It should be understood that the variables listed in Table 6 are suggested at this point based on previous studies which have demonstrated that they are appropriate. Testing with the regression model is required in order to determine those ultimately useful to this study. Additional variables will be introduced should experiments suggest them. A very brief rationale for the expected sign and the importance of the variables is in order. If the difference in price between what publishers charge libraries versus individuals represents price discrimination, then a variable for the individual price (IPRICE) will be a significant predictor of price to institutions (LPRICE). As the individual experiences a rise in price, substitution of access to the library will take place. That is, higher individual prices will shift users toward the library thus raising demand for library subscriptions which will pull institutional prices higher. The sign on this variable is expected to be positive.

One group of variables deals with the issue of price discrimination based on the monopoly power that can be exercised by foreign publishers. Publishers in Great Britain (GBRITAIN), western Europe (EUROPE), and other countries outside the United States (OTHER) may have enough market power to influence price. Therefore these variables will carry a positive sign if there is a sizeable market influence exerted. Some of these publishers will also be concerned with currency exchange risks (RISK), which they will adjust for in prices. However, since they offer discounts through vendors for libraries who prepay subscriptions, this variable will carry a negative sign if the price to individuals captures most of the financial burden of risk adjustment.

It is expected that commercial publishers price discriminate more than their non-profit counterparts. Therefore, in comparison to the commercial residual, associations (ASSOC), government agencies (GOVERN), university presses (UNIVPRESS) and foundations (FOUND) will capture generally lower prices of these non-profit publishers. The signs on all these are expected to be negative.

All the publishers will experience production costs, which can be exposed through variables that control for frequency (FREQ), total pages printed per year (PAGES), peer review (PEERREV) processing/communication expenses and copyright clearance registration expenses (CCCREG), and the presence of graphics, maps, and illustrations (ILLUS), all of which will positively affect price to the extent they are passed along through price discrimination. Circulation (CIRC) will capture the effects of economies of scale, which those publications distributed in larger quantities will experience. Thus this variable is expected to be negative. Similarly, the inclusion of advertising (ADV) will provide additional revenue to that of sales, so this variable is expected to be negative since journals that include ads will have less incentive to extract revenue through sales. New entries into the publishing arena are expected to experience costs for advertising to increase awareness of their products, which will be partially passed on to consumers. Therefore, age (AGE) which is the difference between the current date and the date the journal started will be a negative predictor of price and monopoly power.

Previous studies have developed measures of quality based on rankings of publications compared to each other within a given discipline. Most of these comparisons work from information available from the Institute for Scientific Information. Data acquired from this source showing the impact factor, immediacy index, half-life, total cites, and cites per year will be summarized in one variable to capture quality (QUALITY) of journals. This variable is expected to be positive with regard to both price and monopoly power.

The prices of journals across disciplines may be driven by different factors. In general, prices are higher in the sciences and technical areas and lower in the humanities. This is understandable when we consider the market for science versus humanities outside the academe, whereas scientific publications are used heavily in corporate research by pharmaceutical firms and other industries highly dependent on research. As a result two additional dummies are included in the model to segment the specification along the discipline lines. HUMAN and SOCSCI will control for differences in price among the humanities and social sciences as compared to the residual category of science. These variables are expected to be negative and strong predictors of price.

Table 7. List of Variables.
Dependent variable

LPRICE	The price for library subscriptions.
MPOWER1	Monopoly power as represented by LPRICE minus IPRICE.
MPOWER2	Monopoly power as represented by the index: (LPRICE - IPRICE) / LPRICE
Independent variables	
IPRICE	Price for for individuals. (+, number)
GBRITAIN	1 if the journal published in Great Britain, 0 otherwise. (-, dummy variable)
EUROPE	1 if the journal published in Europe, 0 otherwise. (-, dummy variable)
OTHER	1 if the journal published outside US, Canada, Europe or Great Britain, 0 otherwise.(-, dummy variable)
RISK	Standard deviation of the monthly free market exchange rate between the currency of the home country of a foreign publisher to the U.S. dollar.
ASSOC	1 if the journal is published an association, 0 otherwise. (-, dummy variable)
GOVERN	1 if the journal published by a govt agency, 0 otherwise. (-, dummy variable)
FOUNDTN	1 if the journal published by a foundation, 0 otherwise. (-, dummy variable)
UNIVPR	1 if the journal published by a university press, 0 otherwise. (-, dummy
FREQ	The number of issues per year. (+, number)
PAGES	Number of pages printed per year. (+, number)
PEERREV	1 if article submissions are peer reviewed, 0 otherwise. (+, dummy variable)
CCCREG	1 if journal is registered with the CCC, 0 otherwise. (+, dummy variable)
ILLUS	1 if the journal contains graphics or illustrations, 0 otherwise. (+, dummy)
CIRC	The reported number of subscriptions to the journal. (-, number)
ADV	1 if there is commercial advertising in journal, 0 otherwise. (-, dummy variable)
AGE	Current year minus the date the journal first published. (-, number)
QUALITY	Sum of the Institute for Scientific Information citation measures. (+, number).
HUMAN	1 if the journal is in the humanties, 0 otherwise. (-, dummy variable)
SOCSCI	1 if the journal is in the social sciences, 0 otherwise. (-, dummy variable)
ELECTRONIC	1 if available in electronic form, 0 otherwise. (+, dummy variable)

Finally, a dummy variable is included to determine whether availability of each journal electronically (ELECTRONIC) has a positive impact on ability to price discriminate. Since we have predicted that monopoly power will erode in the electronic arena, ELECTRONIC should be statistically significant and a negative predictor of monopoly power. However, to the extent that availability of a journal electronically distinguishes it from print counterparts, there is some expectation that this variable could be positive. This would capture additional price discrimination by publishers who are able to capture lost revenue in the electronic environment.

The data set will be assembled by enhancing the data on subscriptions gathered during the planning project. Most of the additional dataset elements including prices will be acquired from examination of the journals and invoices received by the libraries. Impact and related factors will be acquired from the Institute for Scientific Information. Circulation will be proxied from the number of subscriptions supplied in print by two major journal vendors, FAXON and Ebsco. An alternative measure of circulation will be compiled from a serials bibliography. The rest of the variables were obtained by examination of the print subscriptions retained by the libraries or from a serials bibliography.

Conclusion

There may be other ways to attack the problem of price inflation of scholarly periodicals. Some hope arises from the production cost differences between print and electronic periodicals. The marginal cost of each added print copy diminishes steadily from the second to the nth copy, whereas for electronic publications, the marginal cost of the second and subsequent copies is approximately zero. Although distribution is not quite zero for each additional copy, since computer resources can be strained by volume of access, the marginal cost is so close to zero that technical solutions to the problem of unauthorized redistribution for free of pirated copies might provide an incentive for publishers in the electronic domain to distribute equitably the cost of the first copy across all consumers. If the total cost of production of the electronic publications is lower than it would be for printed publication, some publishers may share the savings with consumers. However, there is no certainty that they will, because profit maximizers will continue to be profit maximizers. Therefore, it is appropriate to look for a decoupled solution lying in the hands of consumers.

In the meantime, the outcomes of this research project will include a test of the benefits of consortial access versus ownership. In addition, earlier work on price discrimination will be extended with this cross-discipline study to determine whether electronic telecommunications offers hope of relief from monopoly power of publishers.

REFERENCES:

[1] Richard M. Dougherty, "A 'Factory' for Scholarly Journals." *Chronicle of Higher Education* 38/41 (June 17, 1992): b1-b2.

[2] Bert R.Boyce, "Meeting the Serials Cost Problem: a Supply Side Proposal." *American Libraries* 24/3 (March 1993): 272-273.

[3] Steven Harnad, "Post-Gutenberg Galaxy:?The Fourth Revolution in the Means of Production of Knowledge." *Public Access Computer Systems Review* 2/1 (1991): 39-53.

[4] Andrew M.Odlyzko, "Tragic Loss or Good Riddance? The Impending Demise Of Traditional Scholarly Journals." *International Journal of Man-Machine Studies* 42/1 (January 1995):71-122.

[5] Frank Quinn, "A Role for Libraries in Electronic Publication." *Serials Review* 21/1 (1995): 27-30.

[6] Charles A. Schwartz, "Scholarly Communication as a Loosely Structured System: Reassessing Prospects for Structural Reform. *College and Research LIbraries* 55/2 (March 1994): 101-117.

[7] Anthony W. Cummings, Marcia L. Witte, William G. Bowen, Laura O. Lazarus, and Richard H. Ekman, *University Libraries and Scholarly Communication: Study Prepared for the Andrew W. Mellon Foundation*. Washington, DC: Association of Research Libraries for the Andrew W. Mellon Foundation, 1992.

[8] *Newsletter on Serials Pricing Issues*. Edited by Marcia Tuttle. Chapel Hill: University of North Carolina, 1989-1997. Available electronically by subscribing to PRICES from listproc@unc.edu.

[9] Deana Astle and Charles Hamaker, "Journal publishing: pricing and structural issues in the 1930s and the 1980s. *Advances in Serials Management* 2 (1988):1-36

[10] Charles Hamaker and Deana Astle. "Recent Price Patterns in British Journal Pricing." *Library Acquisitions: Practice and Theory* 8 (1984) 225-232.

[11] Deana Astle and Charles Hamaker, "Pricing by Geography: British Journal Pricing 1986 Including Developments in Other Countries." *Library Acquisition s: Practice and Theory* 10 (1986) 165-181.

[12] David Lewis, "Economics of the Scholarly Journal." *College and Research Libraries* 50/6 (November 1989): 674-688.

[13] H. Craig. Peterson, "Variations in Journal Prices: A Statistical Analysis" *Serials LIbrarian* 17/1&2 (1989): 1-9.

[14] Bruce Kingma and Philip Eppard, "Journal Price Escalation and the Market for Information: The Librarians' Solution." *College and Research Libraries* 53/6 (November 1992): 523-535.

[15] Michael A. Stoller, Robert Christopherson,,and , Michael Miranda, "The Economics of Professional Journal Pricing." *College and Research Libraries* 57/1 (January 1996): 9-21.

[16] Henry H. Barshall, "The Cost Effectiveness of Physics Journals." *Physics Today* 41 (July 1988): 56-59.

[17] H. Craig. Peterson, "The Economics Of Economics Journals: A Statistical Analysis of Pricing Practices by Publishers" *College and Research Libraries* 53 (March 1992): 176-181.

[18] John O. Christensen, "Do We Know What We are Paying For? A Comparison of Journal Subscription Costs." *Serials Review* 19/2 (Summer 1993): 39-61.

[19] Edward A. Dyl, "A Note on Price Discrimination by Academic Journals." *Library Quarterly* 53/2 (1983): 161-169.

[20] Patrick Joyce, and Thomas E. Merz, "Price Discrimination in Academic Journals." *Library Quarterly* 55/3 (1985): 273-283.

[21] Patrick Joyce, "Price Discrimination in 'Top' Scientific Journals." *Applied Economics* 22/8 (1990): 1127-1135.

[22] George A. Chressanthis and June D. Chressanthis, "A General Econometric Model of the Determinants of Library Subscription Prices of Scholarly Journals: The Role of Exchange Rate Risk and Other Factors." *Library Quarterly* 64/3 (1994): 270-293.

[23] George A. Chressanthis and June D. Chressanthis, "The Relationship Between Manuscript Submission Fees and Journal Quality." *Serials Librarian* 24/1 (1993): 71-85.

[24] Roger Noll and W. Edward Steinmueller. "An Economic Analysis of Scientific Journal Prices: Preliminary Results" *Serials Review* 18 (Spring/Summer 1992): 32-37.

[25] George A. Chressanthis and June D. Chressanthis, "The Determinants of Library Subscription Prices of the Top-Ranked Economics Journals: An Econometric Analysis" *Journal of Economic Education* 25/4 (Fall 1994): 367-382.

[26] S. J. Liebowitz, "Copying and Indirect Appropriability: Photocopying of Journals." *Journal of Political Economy* 93/5 (1985): 945-957.

[27] For a fuller exploration of the issues that prescribe the study reported here see: Richard W. Meyer, "Monopoly Power and Electronic Journals." *Library Quarterly* 67/4 (October 1997): In press.

[28] Edward Chamberlin, *The Theory of Monopolistic Competition.*Cambridge: Harvard University Press, 1935.

[29] Jan Keppler, *Monopolistic Competition Theory: Origins, Results, and Implications.* Baltimore: Johns Hopkins University Press, 1994.

[30] *Directory of Electronic Journals, Newsletters and Academic Discussion List.* Washington,

DC: Association of Research Libraries, 1991-1996.

[31] Steve Hitchcock, Leslie Carr and Wendy Hall, "A Survey of STM Online Journals 1990-1995: the Calm Before the Storm." In: D. Mogge, editor, *Directory of Electronic Journals, Newsletters and Academic Discussion List.* 6th Edition. Washington, DC: Association of Research Libraries, 1996. Pg. 7-32.

[32] For a copy of the contents of EJournal send: get EJRNL Contents, to listserv@albany.edu.

[33] Stephen P. Harter, "The Impact of Electronic Journals on Scholarly Communication: A Citation Analysis." The Public-Access Computer Systems Review 7, no. 5 (1996): 5-34. [HTML file] URL: http://info.lib.uh.edu/pr/v7/n5/hart7n5.html.

[34] *Adonis* is a product of Adonis USA, 238 Main St., 5th Floor, Cambridge, MA; a wholly owned subsidiary of Elsevier.

[35] Richard W. Meyer, "Management, Cost and Behavioral Issues with Locally Mounted Databases." *Information Technology and Libraries* 9/3 (September 1990): 226-241.

[36] Personal conversation with Karen Hunter, Vice President of Elsevier, 15 February 1997.

[37] George A. Chressanthis and June D. Chressanthis, "Publisher monopoly power and third-degree price discrimination of scholarly journals" *Technical Services Quarterly* 11/2 (1993): 13-36.

[38] This theory is based on the classic work: Abba Lerner, "The concept of monopoly and the measurement of monopoly power" *Review of Economic Studies* (June 1934): 157-175.

[39] John R. Hayes, "The Internet's First Victim?" *Forbes* 156/14 (December 18, 1995): 200-201.

[40] The Associated Colleges of the South includes: Birmingham Southern, Centenary, Centre, Furman, Hendrix, Millsaps, Morehouse (Atlanta University Center), Rhodes, University of Richmond, Rollins, Southwestern, University of the South and Trinity.

[41] Two variations in the dependent variable will be used: the net difference and the index of monopoly power after the work of Lerner.

[42] George A. Chressanthis and June D. Chressanthis. "Publisher monopoly power and third-degree price discrimination of scholarly journals" *Technical Services Quarterly* 11/2 (1993): 13-36.

……

For additional information about the conference, or The Andrew W. Mellon Foundation's scholarly communication initiatives, please contact Richard Ekman. For additional information about ARL or this web site contact Patricia Brennan, ARL Program Officer at (202) 296-2296.

Return to Office of Scholarly Communication Home Page

ARL Home ARL Scholarly Communication and Technology Home Page

© Association of Research Libraries, Washington, DC

Web Design by Angelo F. Cruz
Maintained by ARL Web Administrator
Last Modified: August 29, 1997

Scholarly Communication and Technology

Conference Organized by The Andrew W. Mellon Foundation

at Emory University
April 24-25, 1997

Copyright © of the papers on this site are held by the individual authors or The Andrew W. Mellon Foundation. Permission is granted to reproduce and distribute copies of these works for nonprofit educational or library purposes, provided that the author, source, and copyright notice are included on each copy. For commercial use, please contact **Richard Ekman** at the The Andrew W. Mellon Foundation.

Session #7 Multi-Institutional Cooperation

A New Consortial Model for Building Digital Libraries

Raymond K. Neff
Vice President for Information Services
Director of University Libraries
Case Western Reserve University

rkn@po.cwru.edu

The libraries in America's research universities are being systematically depopulated of current subscriptions to scholarly journals. Annual increases in subscription costs are consistently outpacing the growth in library budgets. This has become a chronic problem for academic libraries which collect in the fields of science, engineering, and medicine, and by now the problem is well recognized (Cummings, 1992). At Case Western Reserve University, we have built a novel digital library distribution system and focused on our collections in the chemical sciences to investigate a new approach to solving a significant portion of this problem. By collaborating with another research library which has a strong chemical sciences collection, we have developed a methodology to control costs of scholarly journals and have planted the seeds of a new consortial model for building digital libraries. This paper summaries our progress to date and indicates areas in which we are continuing our research and development.

For research libraries in academia, providing sufficient scholarly information resources in the chemical sciences represents a large budgetary item. For our purposes, the task of providing high-quality library services to scholars in the chemical sciences is similar to providing services in other sciences, engineering, and medicine; if we solve the problem in the limited domain of the chemical sciences, one can reasonably extrapolate our results to these other fields. Thus, research libraries whose mission it is to provide a high level of coverage for scholarly publications in the chemical sciences are the focus of this project, although we believe that the principles and practices employed in this project are extensible to the serial collections of other disciplines.

A consortium depends on having its members operating with common missions, visions, strategies, and implementations. We adopted the tactics of developing a consortial model by having two neighboring libraries collaborate in the initial project. The University of Akron (UA) and Case Western Reserve University (CWRU) both have academic programs in the chemical sciences which are nationally ranked, and the two universities are fewer than thirty miles apart. It was no surprise to find that both universities have library collections in the chemical sciences which are of high quality and nearly exhaustive in their coverage of scholarly journals. To quantify the correlation between these two collections we counted the number of journals which both collected and found the common set to be 76% in number and 92% in cost. The implications of the overlap in collecting patterns is plain; if both libraries collected only one copy of each journal, with the exception of the most used journals, approximately half of the cost of these subscriptions could be saved. For these two libraries, the cost savings is potentially $400,000 per year. This seemed like a goal worth pursuing, but to do so would require building a new type of information distribution system.

The reason scholarly libraries collect duplicative journals is that students and faculty want to be able to use these materials by going to the library and looking up a particular volume or by browsing the current issues of journals in their field. Eliminating a complete set of the journals at all but one of our consortial libraries would deprive local users of this walk-up-and-read service. We asked ourselves if it would be possible to construct a virtual version of the paper-based journal collection which would be simultaneously present at each consortium member institution, allowing any scholar to consult the collection at will even though only one copy of the paper journal was on the shelf. The approach we adopted was to build a digital delivery system that would provide to a scholar on the campus of a consortial member institution, on a demand basis, either a soft or hard copy of any article for which a subscription to the journal was held by a consortial member library. Thus, according to this vision, the use of information technology would make it possible to collect one set of journals among the consortium members and to have them simultaneously available at all institutions. Although the cost of building the new digital distribution system is substantial, it was considered as an experiment worth undertaking. The generous support of The Andrew W. Mellon Foundation is being used to cover approximately one-half of the costs for the construction and operation of the digital distribution system, with Case Western Reserve University covering the remainder. The University of Akron Library has contributed its expertise and use of its chemical sciences collections to the project.

It also seemed necessary to us to want to invite the cooperation of journal publishers in a project of this kind. To make a digital delivery system practical would require having the rights to store the intellectual property in a computer system, and when we started this project, no consortium member had such rights. Further, it was both the on-going publications and the "back files" which would be needed so that complete "runs" of each serial could be constructed in digital form. The publishers could work out agreements with the consortium to provide their scholarly publications for inclusion in a digital storage system which would be connected to our network-based transmission system, and thus, their cooperation would become essential. The chemical sciences are disciplines in which previous work with electronic libraries had been started. The TULIP Project of Elsevier Science (TULIP, 1996) and the CORE Project of Cornell University, the American Chemical Society, Bellcore, Chemical Abstracts, and OCLC were known to us, and we certainly wanted to benefit from their experiences. Publications of Elsevier Science, the American Chemical Society, and others including Springer-Verlag, the Academic Press, and John Wiley & Sons were central to our proposed project because of the importance of their journal titles to the chemical sciences disciplines.

We understood from the beginning of this effort that we would want to monitor the performance of the digital delivery system under realistic usage scenarios. The implementation of our delivery system has built into it extensive data collection facilities for monitoring what users actually do. The system is also sensitive to concerns of privacy in that it collects no items of performance information which may be used to identify unambiguously any particular user.

Given the existence of extensive campus networks at both CWRU and UA and substantial internetworking among the academic institutions in northeastern Ohio, there was sufficient infrastructure already in place to allow the construction and operation of an intra- and intercampus digital delivery system. Such a digital delivery has now been built and made operational. The essential aspects of the digital delivery system will now be described.

A Digital Delivery System

The roots of the electronic library are found in landmark papers by Bush (1945) and Kemeny (1962). Most interestingly, Kemeny foreshadowed what the prospective scholarly users of our digital library told us as their requirement that they be able to see each page of a scholarly article preserved in its graphical integrity. That is, the electronic image of each page layout needed to look like it did when originally published on paper. The system we have developed uses the ACROBAT^{R} page description language to accomplish this objective.

Because finding aids and indices for specialized publications are too limiting, users also have the requirement that the article's text be searchable with limited or unlimited discipline-specific thesauri. Our system complements the page images with an optical character-recognition (OCR) scanning of the complete text of each article. In this way, the user may enter words and phrases the presence of which in an article would constitute a "hit" for the scholar.

One of the most critical design goals for our project was the development of a scanning subsystem that would be easily reproducible and cost efficient to set up and operate in each consortium member. Not only did the equipment need to be readily available, but it had to be adaptable to a variety of work-flow and staff work patterns in many different libraries. Our initial design has been successfully tailored to the needs of both the CWRU libraries and the Library at the University of Akron. Our approach to the sharing of paper-based collections is to use a scanning device to copy the page images of the original into a digital format which may be readily transmitted across our existing telecommunications infrastructure. In addition, the digital version of the paper original may be stored for subsequent retrieval. Thus, repeated viewing of the same work would necessitate only a one-time transformation of format. This is both an advantage in achieving faster response times for scholars but promotes the development and use of quality control methods. The scanning equipment we have used in this project is the Minolta PS-3000 Digital Planetary Scanner with the Epic 3000 Software Subsystem. The principal advantage of this scanner is that bound serials may be scanned without damaging the volume and without compromising the resulting page images; in fact, the original journal collection remains intact and accessible to scholars throughout the project. This device is also sufficiently fast that a trained operator, including students, may scan over 800 pages per average workday. For a student worker making $7.00 per hour, the person-cost of scanning is under $0.07 per page; the cost of conversion to searchable text adds $0.01 per page. Thus, each consortium member would be expected to make a reasonable investment in equipment, training, and personnel. Appendix D gives more details regarding the scanning processes and workflow. Appendix E gives a technical justification for a digitization standard for the consortium.

The target equipment for viewing an electronic journal was taken to be a common PC-compatible computer workstation, hereafter referred to as a client. This client is also the user platform for the on-line library catalog systems found on our campuses, as well as the growing collections of CD-ROM-based information products. Appendix C gives the specification of the workstation standards for this project. The implications for use of readily available equipment is that the client platform for our project would also work outside of the library - in fact, wherever a user wanted to work. Therefore, by selecting the platform we did, we extended the project to encompass a full campus-wide delivery system. Because our consortium involves multiple campuses (two at the outset), the delivery system is general purpose in its availability as an access facility.

Just as we had within the classical research library a place to store paper-based journals, we needed to specify a place to storage the digital copies. In technical parlance, this storage facility is called a server. To give us the greatest possible flexibility in developing the project, we decided to form the server out of two interlinked computer systems, a standard IBM System 390 with the OS/390 Open Edition version as the operating system and a standard IBM RS/6000 System with the AIX version of the UNIX operating system. Both of these components may be incrementally grown as the project's server requirements increase. Both systems are relatively commonplace at academic sites, although only one system pair is needed in this project, and to provide for both reliability and load leveling, it is likely that eventually two pairs of systems would be needed for an effort on the national scale.

The campus-wide networks on both our campuses and the state-wide network which connects to them uses the standards-based TCP/IP protocols. Thus, any connected client workstation which follows our minimum standards will be able to use the digital delivery system being constructed. Because the key to minimizing the operating costs within a consortium is interoperability and standardization of equipment, we have adopted a series of standards for this project; they are given in Appendices B and C. The minimum transmission speed on the CWRU campus is ten million bits-per-second (M bps) to each client workstation and a minimum of 155 M bps on each backbone link. The principal document repository is on the IBM System 390 which uses a 155 M bps ATM (asynchronous transfer mode) connection to the campus backbone. The linkage to the University of Akron is by way of the state-wide network where the principal backbone connection from CWRU is also operating at 155 M bps, and the linkage from the UA to the state-wide network is at 3 M bps. The on-campus linkage for UA is also a minimum of 10 M bps to each client workstation within the chemical sciences scholarly community and to client workstations in the UA University Library.

One of the most significant problems in placing intellectual property in a networked environment is that with a few clicks of a mouse thousands of copies of the original work can be distributed at virtually zero marginal cost, and the owner is generally deprived of expected royalty revenue. Since we recognized this problem some years ago and we realized that solutions outside of the network itself were unlikely to be either permanent or satisfactory to all parties (e.g., author, owner, publisher, distributor, user), we embarked on the creation of a software subsystem now known as Rights Manager^{TM}. With our RM system, we can control the dissemination of network-based intellectual property subject to each stakeholder receiving his due. Appendix A gives a fuller description of the RM system.

The key to understanding our approach to intellectual property management is that we expect that each scholarly work will be disseminated according to a comprehensive contractual agreement. Publishers may use master agreements to cover a set of titles. Further, we do not expect that there will be only one interpretation of concepts such as "fair use," and our Right Manager system makes provision for arbitrarily different operational definitions of fair use, so that specific contractual agreements can be "enforced" within the delivery system.

A New Consortial Model

The library world has productively used various consortial models for over thirty years, but until now, there has not been a successful model for building a digital library. One of the missing pieces in the consortial jigsaw puzzle has been a technical model which is both comprehensive and reproducible in a variety of library contexts. To begin our approach to a new consortial model, we developed a complete technical system for building and operating a digital library. Building such a system is no small achievement. Similar efforts have been undertaken with the Elsevier Science TULIP Project and the JSTOR project.

The primary desiderata for a new consortial model are as follows:

- Any research library can participate using agreed upon and accepted standards.
- Many research libraries each contribute relatively small amounts of labor by scanning a small, controlled number of journal issues. Scanning is both systematic and based on a request for an individual article.
- Use of readily available off-the-shelf equipment.
- Intellectual property is made available through licensing and controlled by the Rights Manager software system.
- Publishers grant rights to libraries to scan and store intellectual property retrospectively (i.e., already purchased materials) in exchange for the right to license use of the digital formats to other users. Libraries provide publishers with digital copies of scholarly journals for their own use, thus enabling publishers to enrich their own electronic libraries.

A Payments System for the Consortium

It is unrealistic to assume that all use of a future digital library will be without any charging mechanisms even though the research library of today charges for little except for photocopying and user fines. This is not to assume that the library user is charged for each use although that would be possible. More likely it would be the library which would pay on behalf of the members of the scholarly community (i.e., student, professor, researcher) it supports. According to our proposed consortial model, libraries would be charged for use of the digital library according to the total pages "read" in any given user session. It could be easily worked out such that users who consult the digital library on the premises of the campus library would not be charged themselves, but if they used the digital library from another campus location or from off-campus through a network, that they would pay a per-page charge analogous to the cost of photocopying. A system of charging could include categorization by type of user, and the RM system provides for a wide variety of charging models, including the making of distinctions of usage in soft copy format, hard copy format, and downloading of a work in whole or in part. Protecting the rights of the owner is an especially interesting problem when the entire work is downloaded in a digital format. Both visible and invisible watermarking are techniques with which we have experience for protecting rights in the case of downloading an entire work.

We also have in mind that libraries which provide input via scanning to the decentralized, digital library would receive a credit for each page scanned. It is clear that the value of the digital library to the end user will increase as higher degrees of completeness in digitized holdings is achieved. Therefore, the credit system to originating libraries should recognize this and reward these libraries according to a formula that charges and credits with a relative credit-to-charging ratio of perhaps in the neighborhood ten to one; that is, an originating library might receive a credit for scanning equal to a charge for ten soft copy reads.

The charge-and-credit system for our new consortial model is analogous to that used for the highly successful Online Computer Library Center's cataloging system. Member libraries within OCLC contribute original cataloging entries in the form of MARC records for the OCLC database as well as draw down a copy of a holding's data to fill in entries for their own catalog systems. The system of charging for "downloads" and crediting for "uploads" is repeated in our consortial model for retrospective full-text journal articles. Just as original cataloging is at the heart of OCLC, original scanning is at the heart of our new consortial model for building the library of the future.

Data Collection

One of the most important aspects of this project is that we have instrumented the entire software system which underlies the project with data collection points. In this way we can find out through actual usage by faculty, students, and research staff what aspects of the system are good and which need more work and thought. Over the past decade many people have speculated about how the digital library might be made to work for the betterment of scholarly communications. The system described in this paper is one of the most comprehensive attempts yet to have experience benefit visioning.

To appreciate the detailed data being collected by the project, we will describe the various types of data that the RM system captures. Many types of transactions occur between the RM client and the server software throughout a user session. The server software record these transactions to permit detailed analysis of usage patterns. A typical user session generates the following transactions between client and server.

- User requests an article (usually from a Web browser)

  If the user is starting a new session, the RM system downloads and launches the appropriate viewer which will process only encrypted transactions. In the case of Adobe Acrobat, the system downloads a plug-in. The following transactions take place with the server:

  1a. Authenticate the viewer (i.e., ensure we are using a secure viewer).

  1b. Get permissions (i.e., obtain a set of user permissions, if any. If it is a new session, the user is set by default to be the general-purpose category of PUBLIC).

  1c. Get Article (download the requested article. If step b returns no permissions, this transaction does not occur. The user must sign on and request the article again).

- User signs on

  If the general user has no permissions, s/he must log on. Following a successful logon, transactions 1b and 1c must be repeated. Transactions during sign-on include:

  2a. Sign On

- Article is displayed on screen

  Before an article is displayed on the screen, the viewer enters an step-by-step RM process or protocol wherein a single reporting command is sent to the server several times with different state flags and use types. RM events are processed similarly for all supported functions, including display, print, excerpt, and download. The transactions include:

  3a. Report Use BEGIN (Just before the article is displayed).

  3b. Report Use ABORT (Sent in the event that a technical problem prevents display of the article (such as out of memory, etc.)).

  3c. Report Use DECLINE (Sent if the user declines display of the article after seeing the cost).

  3d. Report Use COMMIT (Just after the article is displayed).

  3e. Report Use END (Sent when the user dismisses the article from the screen by closing the article window).

- Users closes viewer

  When a user closes a viewer, an end-of-session process occurs which sends transaction (3e) for all open articles. Also a close viewer transaction is sent which immediately expires the viewer so it may not be used again.

  4a. Close Viewer

The basic data being collected for every command (with the exception of 1a) and being sent to the server for later analysis includes the following:

- Date/Time
- Viewer ID
- User ID (even if it is PUBLIC)
- IP Address of request

These primary data may be used to derive additional data: Transaction (1b) may be effectively used to log unsuccessful access attempts, including failure reasons. The time interval between transactions (3a) and (3e) may be used to measure the duration that an article is on the screen. The basic data collection module in the RM system is quite general and may be used to collect other information and derive other measures of system usage.

Conclusions

A digital distribution system for storing and accessing scholarly communications has been constructed and installed on the campuses of Case Western Reserve University and the University of Akron. This low-cost system can be extended to other institutions with similar requirements because the system components, together with the way they have been integrated, were chosen to facilitate the diffusion of these technologies. This distribution system successfully separates ownership of library materials from access to them.

The most interesting aspect of the new digital distribution system is that it can be the basis for libraries to form consortia which can share highly specialized materials, rather than duplicating them in parallel, redundant collections. When a consortium can share a single subscription to a highly specialized journal, then we have the basis for reducing the total cost of library materials because we can eliminate duplicative subscriptions. We believe that the future of academic libraries points to the maintenance of a basic core collection, the selective acquisition of specialty materials, and the sharing across telecommunications networks of standard scholarly works. The consortial model which we have built and tested is one way to accomplish this goal.

Our approach is contrasted with the common behavior of building up ever larger collections of standard works, so that over time, academic libraries begin to look ever more alike in their collecting habits and offer almost duplicative services and require ever larger budgets. This project is attempting to find another path.

The effects of the new consortial model for building digital libraries are not confined to the domain of technology. During the period when the new digital distribution system was being constructed, an agency of the Ohio Board of Regents called OhioLINK commenced an overlapping experiment with Elsevier Science. According to this recently signed agreement, all of Elsevier Science's eleven-hundred-plus electronic journals will be available for access and use on all of the 55 campuses of OhioLINK member institutions, including CWRU and the University of Akron. The cost of the entire collection of electronic journals for each university for 1997 was set by the OhioLINK contract to be approximately 5.5% greater than the institution's Elsevier Science expenditure level for 1996 subscriptions regardless of the particular subset these subscriptions represented; there is a further 5.5% price increase set to take effect in 1998. Further, the agreement between OhioLINK and Elsevier constrains the member institutions to pay for this comprehensive access even if they cancel a journal subscription. Notably, there is an optional payment discount of 10% when an existing journal subscription (in a paper format) is limited to electronic delivery only (eliminating the delivery of a paper version). Thus, electronic versions of the Elsevier journals which are part of our chemical sciences digital library will be available at both institutions regardless of the existence of our consortium; pooling collections according to our consortial model would be a useless exercise from a financial point of view.

Other publishers are also working with our consortium of institutions to offer digital products. During spring 1997, CWRU and the University of Akron entered into an agreement with Springer-Verlag to evaluate their offering of fifty or so electronic journals, some of which overlapped with our chemical sciences collection. In 1996, OhioLINK also worked out an agreement on behalf of its member institutions with Academic Press to offer their collection of approximately 175 electronic journals, many of which were in our chemical sciences collections. Significantly, the OhioLINK contract with Academic Press facilitated the development of our digital library because it included a provision covering the scanning and storage of retrospective collections (i.e., "backfiles") of their journals which we had originally acquired by subscription. A similar agreement covering backfiles of Elsevier journals is currently under negotiation. During the development of this project, we had numerous contacts with the American Chemical Society with the objective of including their publications in our digital library. Indeed, the outline of an agreement with them was discussed. As the time came to render the agreement in writing, they withdrew and later disavowed any interest in a contract with the consortium. At the present time, discussions are being held with other significant chemical science publishers about being included in our consortial library. This is clearly a dynamic period in journal publishing and each of the societal and commercial publishers sees much at stake. While we in universities try to make sense of both technology and information service to our scholarly communities, the publishers are each trying to chart their own course both competitively and strategically while improvements in information technology continually raise the "ante" for continuing to stay in the "game."

Over the past decade several interesting experiments have been conducted to test different ideas for developing digital libraries, and more are under way. With many differing ideas and visions, an empirical approach is a sound way to make progress from this point forward. Our consortium model with its many explicit standards and integrated technology seems to us to be an experiment worth continuing. During the next few years it will surely develop a base of performance data which should provide insights for the future. In this way, experience will benefit visioning.

REFERENCES:

Borghuis, M., Brinckman, H., Fischer, A., Hunter, K., van der Loo, E., Mors, R., Mostert, P., and Zilstra, J.: TULIP Final Report: *T*he *U*niversity *Li*censing *P*rogram. New York: Elsevier Science, 1996.

Bush, V.: "As We May Think" The Atlantic Monthly, 176, 101-108, 1945.

Cummings, A.M., Witte, M.L., Bowen, W.G., Lazarus, L.O., Ekman, R.H.: University Libraries and Scholarly Communication: A Study Prepared for The Andrew W. Mellon Foundation. The Association of Research Libraries, 1992.

Fleischhauer, C. and Erway, R.L.: "Reproduction-Quality Issues in a Digital-Library System: Observations on the Reproduction of Various Library and Archival Material Formats for Access and Preservation." An American Memory White Paper, Washington, D.C.: Library of Congress, 1992.

Kemeny, J.G.: "A Library for 2000 A.D." in Greenberger, M. (Ed.), Computers and the World of the Future. Cambridge, MA.: The M.I.T. Press, 1962.

Appendix A: Rights Manager^{TM}

Case Western Reserve University has developed a rights management system (called Rights Manager^{TM}) for controlling the distribution of digitally formatted intellectual property in a networked environment. This appendix is a high-level description of the system.

CWRU has been working for the past seven years to address various problems in building a digital library. During this period, it has collaborated on a variety of projects involving multimedia authoring and presentation software systems; however, its primary objective has been the development of a client server-based content delivery system that manages intellectual property distribution for digitally formatted content (e.g., text, images, audio, video, and animations).

Rights Manager is a working system that encodes license agreement information for intellectual property at a server and distributes the intellectual property to authorized users over the Internet or a campus-wide Intranet along with a Rights Manager-compliant browser. The Rights Manager handles a variety of license agreement types, including public domain, site licensed, controlled simultaneous accesses, and pay-per-use. Rights Manager also manages the functionality available to a client according to the terms of the license agreement; this is accomplished by use of a special browser that enforces the license's terms and which permits or denies client actions such as save, print, display, copy, etc. Access to a particular item of intellectual property, with or without additional functionality, may be made available at no charge, with an overhead charge, or at a royalty plus overhead charge to the client. Rights Manager has been designed to accommodate sufficient flexibility in capturing wide degrees of arbitrariness in charging rules and policies.

The Rights Manager is intended for use by individuals and organizations who function as purveyors of information (publishers, on-line service providers, campus libraries, etc.). The system is capable of managing a wide variety of agreements from an unlimited number of content providers. Rights Manager also permits customization of licensing terms so that individual users or user classes may be defined and given unique access privileges to restricted sets of materials. A relatively common example of this for CWRU would be an agreement to provide (a) view-only capabilities to an electronic journal accessed by an anonymous user located in the library, (b) display/print/copy access to all on-campus students enrolled in a course for which the digital textbook has been adopted, and (c) full access to faculty for both student- and instructor-versions of digital versions of supplementary textbook materials.

Fundamental to the implementation of Rights Manager are the creation and maintenance of distribution rights, permissions and license agreement databases. These databases express the terms and conditions under which the content purveyor distributes materials to its end-users. Relevant features of Rights Manager include:

- a high degree of granularity for publisher-defined content
- central or distributed management of rights, permissions and licensing databases
- multiple agreement types (e.g., site licensing, limited site licensing and pay-per-use)
- content packaging where rights and permission data are combined with digital format content elements for managed presentation by Web browser "plug-in" modules or helper applications.

Rights Manager maintains a comprehensive set of distribution rights, permissions, and charging information. The premise of Rights Manager is that each publication may be viewed as a compound document. A publication under this definition consists of one or more content elements and media types; each element may be individually managed, as may be required, for instance, in an anthology.

Individual content elements may be defined as broadly or narrowly as required (i.e., the granularity of the elements is defined by the publisher); however, for overall efficiency, each content element should represent a significant and measurable unit of material. Figures, tables, illustrations, and text sections may reasonably be defined as content elements.

To manage the distribution of complete publications or individual content elements, two additional licensing metaphors are implemented. The first of these, a Collection Agreement, is used to specify an agreement between a purveyor and its supplier (e.g., a primary or secondary publisher); this agreement takes the form of a list of publications distributed by the purveyor and the terms and conditions under which these publications may be issued to end-users (one or more Collection Agreements may be defined and simultaneously managed between the purveyor and a customer).

The second abstraction, a Master Agreement, is used to broadly define the rules and conditions that apply to all Collection Agreements between the purveyor and its content supplier. Only one Master Agreement may be defined between the supplier and the institutional customer. In practice, Rights Manager assumes that the purveyor will enter into licensing agreements with its suppliers for the delivery of digitally formatted content. At the time the first license agreement is executed between a supplier and a purveyor, one or more entries are made into the purveyor's Rights Manager databases to define the Master and Collection Agreements. Optionally, Publication and/or Content-Element usage rules may also be defined. Licensed materials may be distributed from the purveyor's site (or perhaps by an authorized service provider); both the content and associated licensing rules are transferred by the supplier to the purveyor for distributed license and content management.

Depending upon the selected delivery option, individual end-users (e.g., faculty members, students or library patrons) may access either a remote server or a local institutional repository to search and request delivery of licensed publications. Depending upon the agreement(s) between the owner and the purveyor, individual users are assigned access rights and permissions based upon user-IDs, network addresses, or both.

Network or Internet Protocol addresses are used to limit distribution by physical location (e.g., to users accessing the materials from a library, a computer lab or from a local workstation). User identification may be exploited to create limited site-licensing models or individual user agreements (e.g., distributing publications only to students enrolled in Chemistry 432 or, perhaps, to a specific faculty member).

At each of the four permissioning levels (Master Agreement, Collection Agreement, Publication, and Content-Element), access rules and usage privileges may be defined. In general, the access and usage permissions rules are broadly defined at the Master and Collection Agreement level and are refined or restricted at the Publication and Content-Element levels. For example, a general license agreement rule could be defined to specify that by default all licensed text elements may be printed at a some fixed cost, say 10¢ per page; however, high value or core text sections may be individually identified and assessed higher charges, say 20¢ per page, using publication or content element override rules.

When a request for delivery of materials is received, the content rules are evaluated in a bottom-up manner (e.g., content element rules are evaluated before publication rules which are, in turn, evaluated before license agreement rules, etc.). Access and usage privileges are resolved when the system first recognizes a match between the requester's user-ID (or user category) and/or the network address and the permission rules governing the content. Access to the content is only granted when an applicable set of rules specifically granting access permission to the end-user is found; in the case where two or more rules permit access, the rules most favorable to the end-user are selected. Under this approach, site licenses, limited site licenses, individual licensing, and pay-per-use may be simultaneously specified and managed.

The following use of the Rights Manager rules databases is recommended as an initial guideline for Rights Manager implementation:

> 1) Use Master rules to define the publishing holding company or imprint, the agreement's term (beginning and ending dates), and the general "fair use" guidelines negotiated between a supplier and the purveyor. Because of the current controversy over the definition of "fair use," Rights Manager does not rely upon preprogrammed definitions; rather, the supplier and purveyor may negotiate this definition and create rules as needed. This approach permits "fair use" definitions to be re-defined in response to new standards or regulatory definitions without requiring modifications to Rights Manager itself.
>
> 2) Use Collection Agreement rules to define the term (beginning and ending dates) for specific licensing agreements between the supplier and the purveyor. General access and permission rules by user-ID, user category, network address, and media type would be assigned at this level.
>
> 3) Use Publication rules to impose any user-ID or user category-specific rules (e.g. permissions for students enrolled in a course for which this publication has been selected as the adopted textbook) or to impose exceptions based on the publication's value.
>
> 4) Use Content-Element rules to grant specific end users or user categories access to materials (e.g., define content elements which are supplementary teaching aids for the

instructor) or to impose exceptions based on media type or the value of content elements.

The Rights Manager system does not mandate that licensing agreements exploit user-IDs; however, maximum content protection and flexibility in license agreement specification is achieved when this feature is used. Given that many institutions or consortium customers may not have implemented a robust user authentication system, alternative approaches to uniquely identifying individual users must be considered. While there are a variety of ways in which to address this issue, it is suggested that PIN numbers, assigned by the supplier and distributed by trusted institutional agents at the purveyor's site (e.g., instructors, librarians, bookstore employees or departmental assistants) or embedded within the content be used as the basis for establishing user-IDs and passwords. Using this approach, valid users may enter into registration dialogs to automatically assign user-IDs and passwords in response to a valid PIN "challenge."

While Rights Manager is designed to address all types of multimedia rights, permissions and licensing issues, the current implementation has focused on distribution of traditional print publication media (text and images). Extensions to Rights Manager will be required to address the distribution of full multimedia.

Appendix B: Consortial Standards

MARC

- Enumeration and chronology standards from the serials holding standards of the 853 and 863 fields of MARC
  + Specifies up to 6 levels of enumeration and 4 levels of chronology

  e.g.,

  853 |aVolume|bIssue|i(year)|j(month)
  853 |aVolume|bIssue|cPart|i(year)|j(month)

- Linking from bibliographic records in library catalog via an 856 field
  + URL information appears in subfield "u", anchor text appears in subfield z

  e.g.,

  856 7 |uhttp://beavis.cwru.edu/chemvl|zRetrieve articles from the Chemical Sciences Digital Library

  Would appear as
  Retrieve articles from the Chemical Sciences Digital Library

TIFF

- Most widely used multi-page graphic format
- Support for tagged information ("Copyright", etc.)
- Format is extensible by creating new tags (such as RM rule information, authentication hints, encryption parameters)
- Standard supports multiple kinds of compression

Adobe PDF

- Container for article images
- Page description language
- PDF files are searchable by the Adobe Acrobat browser
- Encryption and security are defined in the standard

SICI (Serial Item and Contribution Identifier)

- SICI Definition (Standards progress, overview, etc.)
- Originally a key part of the indexing structure
- All of the components of the SICI code are stored, so it could be used as a linking mechanism between an article database and the ChemVL Library
- OhioLINK is also very interested in this standard, and is pushing database creators and search engine providers to add SICI number retrieval to citation database and journal article repository systems.
- Future retrieval interfaces into the database: SICI number search form, SICI number search API

  e.g., 0022-2364(199607)121:1<83:TROTCI>2.0.TX;2-I

Appendix C: Equipment Standards for End-Users

Minimum Equipment Required

Hardware: An IBM PC or compatible computer with the following components:

- 80386 processor
- 16MB RAM
- 20MB free disk space
- A video card and display monitor with a resolution of 640 x 480 and 16 colors or shades of gray.

Software:

- Windows^{R} 3.1
- Win32s 1.25
- TCP/IP software suite including a version of Winsock
- Netscape Navigator^{R}2.02
- Adobe Acrobat Exchange^{R} 2.1

Win32s is a software package for Windows 3.1 which is distributed without charge and is available from Microsoft.

The requirement for Adobe Acrobat Exchange, a commercial product which is not distributed without charge, is expected to be relaxed in favor of a requirement for Adobe Acrobat^{R} Reader, a commercial product which is distributed without charge.

The software will also run on newer versions of compatible hardware and/or software.

Recommended Configuration of Equipment

This configuration is recommended for users who will be using the system extensively.

Hardware: A computer with the following components

- Intel Pentium^{R} processor
- 32MB RAM
- 50MB free disk space
- A video card and display monitor with a resolution of 1280 x 1024 and 256 colors or shades of gray.

Software

- Windows NT^{R} 4.0 Workstation
- TCP/IP suite which has been configured for a network connection
- (included in Windows NT)
- Netscape Navigator^{R} 2.02
- Adobe Acrobat Exchange^{R} 2.1

The requirement for Adobe Acrobat Exchange^{R}, a commercial product which is not distributed without charge, is expected to be relaxed in favor of a requirement for Adobe Acrobat^{R} Reader, a commercial product, which is distributed without charge.

Other software options the system has been tested on include:

- IBM OS/2 3.0 Warp Connect^{R} with Win-OS/2
- IBM TCP/IP for Windows 3.1, version 2.1.1
- Windows NT 3.51

Appendix D: Scanning and Workflow

Article Scanning, PDF Conversion and Image Quality Control

The goal of the scan-and-store portion of the project is to develop a complete and tested system of hardware, software and procedures that can be adopted by other members of the consortium with a reasonable investment in equipment, training and personnel. If a system is beyond a consortium member's financial means, it will not be adopted. If a system cannot perform as required, it is a waste of resources.

Our original proposal stressed that all existing scholarly resources, particularly research tools, would remain available to scholars throughout this project. To that end, the scan-and-store process is designed to leave the consortium's existing journal collection intact and accessible.

Scan-and-Store Process Resources

- Scanning workstation, including a computer with sufficient processing and storage capacity, a scanner, and a network connection. Optionally, a second workstation can be used by the scanning supervisor to process the scanned images. The workstation used in this phase of the project includes:

  + Minolta PS-3000 Digital Planetary Scanner
  + Two computers with Pentium 200MHz CPU, 64Mb RAM, 4Gb HD, 21" monitor

  + Windows 3.11 OS (required by other software)
  + Minolta Epic 3000 scanner software
  + Adobe Acrobat Capture, Exchange, and Distiller software
  + Image Alchemy software
  + Network interface cards and TCP/IP software for campus network access

- Scanner operator(s), typically student assistants, with training roughly equivalent to that required for Inter-Library Loan photocopying. Approximately 8 hours of operator labor will be required to process the average 800 pages per day capacity of a single scanning workstation.

- Scanning supervisor, typically a librarian or full-time staff, with training in image quality control, indexing and cataloging, and operation of image processing software. Approximately 3 hours of supervisor labor will be required to process 800 scanned pages per day.

Scan-and-Store Process: Scanner Operator

- Retrieve scan request from system
- Retrieve materials from shelves (enough for two hours of scanning)
- Scan materials and enter basic data into system

  + evaluate size of pages
  + evaluate grayscale/black and white scan mode
  + align material
  + test scan and adjust settings and alignment as necessary
  + scan article
  + log changes and additions to author, title, journal, issue and item data on request form
  + repeat for remaining requested articles

- Transfer scanned image files to Acrobat conversion workstation
- Retrieve next batch of scan requests from system
- Reshelve scanned materials and retrieve next batch of materials

Scan-and-Store Process: Acrobat conversion workstation

- Run Adobe Acrobat Capture to automatically convert sequential scanned image files from single-page TIFF to multi-page Acrobat PDF documents, as they are received from scanner operator
- Retain original TIFF files

Scan-and-Store Process: Scanning Supervisor

- Retrieve request forms for scanned materials
- Open converted PDF files
- Evaluate image quality of converted PDF files

  + scanned article matches request form citation
  + completeness, no clipped margins
  + legibility, especially footnotes and references
  + minimal skewing
  + clarity of grayscale or halftone images
  + appropriate margins, no excessive white space

- Crop fingertips, margin lines, etc., missed by Epic 3000 scanner software

  + retrieve TIFF image file

  + mask unwanted areas
  + re-save TIFF image file
  + repeat PDF conversion
  + evaluate image quality of revised PDF file

- Return unacceptable scans to scanner operator for re-scan or correction
- Evaluate, correct and expand entries in request forms
- Forward corrected PDF files to the database
- Delete TIFF image files from conversion workstation

Notification to and Viewing by User of Availability of Scanned Article

Insertion of the article into the database

- The scanning technician types in the scan request number into a web form.
- The system returns a web form with most of the fields filled in. The technician has an opportunity to correct information from the paging slip before inserting the article into the database.
- The web form contains a "file upload" button that when selected allows the technician to browse the local hard drive for the article PDF file. This file is automatically uploaded to the server when the form is submitted.
- The system inserts the table of contents information into the database and the PDF file to the RightsManager system.

Notification/delivery of article to requester

- E-mail to requester with URL of requested article (in first release)
- No notification (in first release)
- FAX to requester an announcement page with the article URL (proposed future enhancement)
- FAX to requester a copy of the article (proposed future enhancement)

Appendix E: Technical Justification for A Digitization Standard for the Consortium

It is a major premise in the technical underpinnings of the new consortial model that a relatively inexpensive scanner can be located in the major academic libraries of consortium members. After evaluating virtually every scanning device in the market, including some in laboratories under development, we concluded that the 400 dot-per-inch (dpi) scanner from Minolta was fully adequate for the purpose of scanning all the hundreds of chemical sciences journals in which we were interested. Thus, for our consortium, the Minolta 400 dpi scanner was taken to be the digitization standard. The standard which was adopted preserves 100% of the informational content required by our end-users.

More formally, the standard for digitization in the consortium is defined as follows:

The scanner captures 256 levels of gray in a single-pass with a density of 400 dots-per-inch and converts the gray-scale image to black-and-white using threshold and edge-detection algorithms.

We arrived at this standard by considering our fundamental requirements:

- Handle the smallest significant information presented in the source documents of the chemical sciences literature, which is the lower-case e in super- or sub-scripts as occur in footnotes
- Satisfy both legibility and fidelity to the source document
- Minimize scanning artifacts or "noise" from background
- Operate in the range of preservation scanning
- Be affordable by academic and research libraries

The scanning standard adopted by this project was subjected to tests of footnoted information, and 100% of the occurrences of these characters were captured in both image and character modes and recognized for displaying and searching.

At 400 dpi, the Minolta scanner works in the range of preservation quality scanning as defined by researchers at the Library of Congress (Fleischhauer and Erway, 1992).

We were also cautioned about the problems unique to very high resolution scanning where the scanner produces artifacts or "noise" from imperfections in the paper used. It is a happy note that this was not a problem which we have encountered in this project because the paper used by publishers of chemical sciences journals is coated.

When more is less: Images scanned at 600 dpi require larger file sizes than those scanned at 400 dpi. Thus, 600 dpi is less efficient than 400 dpi. Further, in one series of tests which we conducted, a 600 dpi scanner actually produced an image of effectively lower resolution than 400 dpi. It appears that this loss of information occurs when the scanned image is viewed on a computer screen where there is relatively heavy use of anti-aliasing in the display. When viewed with software which permitted zooming-in for looking at details of the scanned image (which is supported by both PDF and TIFF viewers), the 600 dpi anti-aliased image actually had lower resolution than an image produced from the same source document by the 400 dpi Minolta scanner according to our consortium's digitization standard. With the 600 dpi scanner, the only way for the end-user to see the full resolution was to download the image and then print it out. When a comparison was made of the "soft copy" displayed images, the presentation image quality of 600 dpi was unacceptable to our end-users; the 400 dpi image was just right. Thus, our delivery approach is more useful to the scholar who needs to examine fine details on-screen. We conducted some tests by reconstructing the journal page from the scanned image by printing it out on a Xerox DocuTech 6135 (600 dpi). We found that the smallest fonts actually used and fine details of the articles were uniformly excellent. Interestingly, in many of the tests we performed, our faculty colleagues judged the end result by their own "acid test:" how good was the scanned image when printed out in comparison with that produced by a photocopier. For the consortium standard, they were satisfied with the result and pleased with the improvement in quality that the 400 dpi scanner provided in comparison with conventional photocopying of the journal page.

• •

For additional information about the conference, or The Andrew W. Mellon Foundation's scholarly communication initiatives, please contact Richard Ekman. For additional information about ARL or this web site contact Patricia Brennan, ARL Program Officer at (202) 296-2296.

Return to Office of Scholarly Communication Home Page

ARL Home ARL Scholarly Communication and Technology Home Page

© Association of Research Libraries, Washington, DC

Web Design by Angelo F. Cruz
Maintained by ARL Web Administrator
Last Modified: August 29, 1997

Scholarly Communication and Technology

Conference Organized by The Andrew W. Mellon Foundation

at Emory University
April 24-25, 1997

Copyright © of the papers on this site are held by the individual authors or The Andrew W. Mellon Foundation. Permission is granted to reproduce and distribute copies of these works for nonprofit educational or library purposes, provided that the author, source, and copyright notice are included on each copy. For commercial use, please contact **Richard Ekman** at the The Andrew W. Mellon Foundation.

Session #8 Sustaining Change

Information Based Productivuty

Scott Bennett
Univesity Librarian
Yale University

• •

INFORMATION-BASED PRODUCTIVITY

Convenience is a key word in the library lexicon. As service organizations, libraries give high priority to enhancing the convenience of their operations. Readers themselves regularly use the word to describe what they value.[1] By contrast, when NEXIS-LEXIS describes itself as a sponsor of public radio, it emphasizes not convenience but productivity for professionals. Does NEXIS-LEXIS know something that we are missing?

I think so. Talk about productivity is unambiguously grounded in the discourse of economics, whereas talk about convenience rarely is. Quite notably, the Andrew W. Mellon Foundation has self-consciously insisted that its programs in scholarly communication operate within the realm of economics. Foundation President William G. Bowen explains this focus, in speaking of the Foundation's JSTOR project, by observing that "when new technologies evolve, they offer benefits that can be enjoyed either in the form of more output (including opportunities for scholars to do new things or to do existing tasks better) or in the form of cost savings . . . . In universities electronic technologies have almost always led to greater output and rarely to reduced costs . . . . This proclivity for enjoying the fruits of technological change mainly in the form of `more and better' cannot persist. Technological gains must generate at least some cost savings."[2] In its JSTOR project and the other scholarly communication projects it supports, the Foundation calls for attention "to economic realities and to the cost-effectiveness" of different ways of meeting reader needs. The Foundation wishes to promote change that will endure because the changes embody "more effective and less costly ways of doing [the] business" of both libraries and publishers.[3]

Productivity is the underlying measure of such effectiveness, so I want briefly to recall what economists mean by the word and to reflect on the problematic application of productivity measures to higher education. I will then describe a modest project recently undertaken to support one of the most famous of Yale's undergraduate courses. I will conclude with some observations about why the productivity of libraries and of higher education must command our attention.

PRODUCTIVITY

Productivity is one of the most basic measures of economic activity. Comparative productivity figures are used to judge the efficiency with which resources are used, standards of living changed, and wealth created.[4] Productivity is the ratio of what is produced to the resources required to produce it, or the ratio of economic outputs to economic inputs:

$$\text{Productivity} = \frac{\text{Outputs}}{\text{Inputs}}$$

Outputs can be any goods, services, or financial outcomes; inputs are the labor, services, materials, and capital costs incurred in creating the output. If outputs increase faster than inputs, productivity increases. Conversely, if inputs increase faster than outputs, productivity falls. Technological innovation has historically been one of the chief engines of productivity gain.[5]

Useful indicators of productivity require that both inputs and outputs be clearly defined and measured with little ambiguity. Moreover, the process for turning inputs into outputs must be clearly understood. And those processes must be susceptible to management if productivity increases are to be secured. Finally, meaningful quality changes in outputs need to be conceptually neutralized in measuring changes in productivity.

One need only list these conditions for measuring and managing productivity to understand how problematic they are as applied to higher education.[6] To be sure, some of the least meaningful outputs of higher education can be measured, such as the number of credit hours taught or degrees granted. But the outputs that actively prompt people to pursue education--enhanced knowledge, aesthetic cultivation, leadership ability, economic advantage, etc.--are decidedly difficult to measure. And while we know a great deal about effective teaching, the best of classroom inputs remains more an art in the hands of master teachers than a process readily duplicated from person to person. Not surprisingly, we commonly believe that few teaching practices can be consciously managed to increase productivity and are deeply suspicious of calls to do so.

Outside the classroom and seminar, ideas of productivity have greater acceptance. Productive research programs are a condition of promotion and tenure at research universities; and while scholars express uneasiness about counting research productivity, it certainly happens. The ability to generate research dollars and the number of articles and books written undeniably count, along with the intellectual merit of the work. There is little dispute that many other higher education activities are appropriately judged by productivity standards. Some support services, such as the financial management of endowment resources, are subject to systematic and intense productivity analysis. Other academic support activities, including the provision of library services, are expected to be efficient and productive, even where few actual measures of their productivity are taken.[7]

In many cases, discussion of productivity in higher education touches highly sensitive nerves.[8] Faculty, for instance, commonly complain that administration is bloated and unproductive. Concern for the productivity of higher education informs a significant range of the community's journalistic writing and its scholarship.[9] This sensitivity reflects the truly problematic application of productivity measures to much that happens in education and the tension between concerns about productivity and quality. But it also reflects the fact that we are "unable and, on many campuses, unwilling to answer the hard questions about student learning and educational costs" that a mature teaching enterprise is inescapably responsible for answering.[10]

THE SCULLY PROJECT

A modest digital project undertaken last year at Yale offers an opportunity to explore productivity matters. The project aimed at improving the quality of library support and of student learning in one of the most heavily enrolled undergraduate courses at Yale. We wished to do the project as cost-effectively as possible, but initially we gave no other thought to productivity matters. To echo Bowen's words, we wanted to take the fruits of digital technology in the form of more output, as "more and better." But the project provided an opportunity to explore possibilities for cost savings, for reduced inputs. The project, in spite of its modest objectives and scale (or perhaps exactly for those reasons!), became an instructive "natural experiment" in scholarly communication very much like those supported by the Mellon Foundation.

For years, Emeritus Professor Vincent Scully has been teaching his renowned Introduction to the History of Art, from Prehistory to the Renaissance. The course commonly enrolls 500 students, or about 10% of the entire undergraduate student body at Yale. Working with Professor Mary E. Miller, head of the History of Art department, and with Elizabeth Owen and Brian Allen, Head Teaching Fellows with substantial experience in Professor Scully's course, Max Marmor, the head of Yale's Arts Library, and his colleague Christine de Vallet undertook to provide improved library support for this course. Their Scully Project was part of a joint program between the University Library and Information Technology Services at Yale designed to offer targeted support to faculty as they employ digital technologies for teaching, research, and administration. The Scully Project was also our first effort to demonstrate what it could mean to move from film-based to digitally-based systems to support teaching in art history.[11]

The digital material created for Professor Scully's students included:

- An extensive and detailed course syllabus, including general information about the course and requirements for completing it.
- A roster of the 25 Teaching Fellows who help conduct the course, complete with their e-mail addresses, and a schedule of section meetings.
- A list of the four required texts and the six journal articles provided in a course pack.
- A comprehensive list of the works of art discussed in the course, along with detailed information about the artists, dates of creation, media and size, and references to texts that discuss the works.

Useful as this textual material is, it would not meet the course's key information need for images. The Scully Project therefore includes 1,250 images of sculptures, paintings, buildings, vases, and other objects. These images are presented in a Web image browser that is both handsome and easily used, and accompanied by a written guide advising students on study strategies to make the best use of the Web site.[12]

How did the Scully project change student learning? To answer that question, I must first describe how the library used to meet the course's need for study images. The library traditionally selected mounted photographs closely related to, but not necessarily identical to the images used in Professor's Scully's lectures. We hung the photographs in about 480 square feet of study gallery space in the History of Art department. Approximately 200 photographs were available to students for four weeks before the mid-term exam and 400 photographs for four weeks before the final exam. In those exams students are asked to identify images and to comment on them. With 500 students enrolled, and with the photos available in a relatively small space for just over half of the semester, the result was extreme crowding of students primarily engaged in visual memorization. To deal with the obvious imperfections of this arrangement, some of Professor Scully's more entrepreneurial students made video tapes of the mounted photos and sold them for study in the residential colleges. Less resourceful students simply stole the photos from the walls.

The Scully Project employed information technology to do more and better.

- Students can study the slide images Professor Scully actually uses in class, rather than frequently different photographs that are often in black-and-white rather than color and sometimes carry out-dated identifying labels.
- The 1,250 digital images on the Web site include not only those that Professor Scully uses in class, but also other views of the same object and still other images the Teaching Fellows refer to in discussion sessions. Students now have easy access to three times the number of images they could see in the study gallery space. For instance, where before they had one picture of Stonehenge, they now have eight, including a diagram of the site and drawings showing construction methods and details.
- Digital images are available for study throughout the semester, not just before term exams. They are also available at all hours of day and night, consistent with student study habits.
- The digital images are available as a Web site anywhere there is a networked computer at Yale. This includes the residential colleges, where probably three-fourths of undergraduates have their own computers, as well as computing clusters at various locations on campus.

- The images are usually of much better quality than the photographs mounted on the wall; they read to the screen quickly in three different magnifications; and they are particularly effective on 17" and larger monitors.
- The digital images cannot be stolen or defaced. They are always available in exactly the form intended by Professor Scully and his Teaching Fellows.

Student comments on the Scully Projects emphasized the convenience of the Web site. Comments like "convenient, comfortable, detailed all at the push of a button," and "fantastic for studying for exams" were common, as were grateful comments on the 24-hour a day availability of the images and the need not to fight for viewing space in the study gallery. One student told us "it was wonderful. It made my life so much easier." Another student said "it was very, very convenient to have the images available on-line. That way I could study in my own room in small chunks of time instead of having to go to the photo study. I mainly just used the web site to memorize the pictures like a photo study in my room."[13]

Visual memory training is a key element in the study of art history, and the Scully web site was used primarily for memorization. Reports from Teaching Fellows on whether the digital images enhanced student learning varied, and only two of the Fellows had taught the course before and could make comparisons between the photo study space and the Web site. The following statements represent the range of opinion:

- Students "did think it was `cool' to have a web site but [I] can't say they wrote better or learned more due to it."
- "I don't think they learned more, but I do think it [the Web site] helped them learn more easily."
- The head Teaching Fellow for the course reported that student test performance on visual recognition was "greatly enhanced" over her previous experience in the course. Another Teaching Fellow reported that students grasped the course content much earlier in the semester because of the earlier availability of the Web site images.
- One Teaching Fellow expressed an unqualified view that students learned more, wrote better papers, participated in class more effectively, and enjoyed the course more because of the Scully Project.[14]
- Another Teaching Fellow commented, I "wish we had such a thing in my survey days!"

The Web site apparently contributed significantly to at least one key part of Professor Scully's course--that concerned with visual memory training. We accomplished this at reasonable cost. The initial creation of digital images cost about $2.25 an image, while the total cash outlay for creating the Web site was $10,500. We did not track computing costs or the time spent on the project by permanent university staff, but including these costs might well drive the total to about $17,200 and the per image cost to around $14. Using this higher cost figure, one might say we invested $34 for every student enrolled in the course, or $11 per student if one assumes the database remains useful for six years and the course is offered every other year.

This glow of good feeling about reasonable costs, quality products, improved learning, and convenience for readers is often as much as one has to guide decisions on investing in information technology. Last year, however, Yale Professor of Cardiology Carl Jaffe took me up short by describing the criterion by which he judges his noteworthy work in instructional media.[15] For Professor Jaffe, improved products must help solve the cost problem of good education. One must therefore ask whether the Scully Project passes not only the test of educational utility and convenience set by Professor Scully's Teaching Fellows, but also the productivity test set by Professor Jaffe. Does the Scully Project help solve cost problems in higher education? Does it allows us to use university resources more productively?

ACHIEVING INFORMATION-BASED PRODUCTIVITY GAINS

For more than a generation, libraries have been notably successful in improving the productivity of their own operations with digital technology. It is inconceivable that existing staff could manage today's circulation work load if we were using McBee punch cards or--worse yet--typewriter-written circulation cards kept in book-pockets and marked with date stamps attached to the tops of pencils. While libraries have an admirable record of deploying information technology to increase the productivity of their own operations, and while there is more of this to be done, the most important productivity gains in the future will lie elsewhere. The emergence of massive amounts of textual, numeric, spatial, and image information in digital formats, and the delivery of that information through networks, is decisively shifting the question to one of teacher and reader productivity.

What does the Scully Project tell us about library, teacher, and reader productivity? To answer that question, I will comment first on a set of operational issues that includes the use of library staff and Teaching Fellows to select and prepare images for class use; the preservation of the images over time; and the use of space. I will assess the Scully Project both as it was actually deployed, with little impact on the conduct of classroom instruction, and as one might imagine it being deployed as the primary source of images in the classroom. The operations I will describe are more or less under the university's administrative control, and savings achieved in any of them can at least theoretically be pushed to the bottom line or redirected elsewhere. I will also comment on student productivity. This is a much more problematic topic because we can barely imagine controlling or redirecting for productivity purposes any gains readers might achieve.

Productivity gains subject to administrative control

The comparative costs of selecting images and preparing them for instructional use in both the photographic and digital environments are set out in the four tables that follow. These tables are built from a cost model of over three dozen facts, estimates, and assumptions about Professor Scully's course and the library support it requires.[16] Appendix 1 presents the model, with some information obscured to protect confidentiality. I do not explain the details of the cost model[17] here but focus instead on what it tells us. One cautionary word is in order. The cost model generates the numbers given in the tables, but these numbers are probably meaningful only to the nearest $500. In the discussion that follows, I round the numbers accordingly.

The first table compares the cost of library support for Professor Scully's course in its former dependence on photos exhibited in the study gallery and in its present dependence on digital images delivered in a Web site.[18]

TABLE 1, "AS DONE" CONDITION: 1,250 images used primarily for memory training

1st Year and Cumulative 6-Yr Expenses	400 Photos		1,250 Digital Images	
Selection of images	1st Year	6-Year Total	1st Year	6-Year Total
Full-time library staff for photo collection	797	2,392	6,200	7,440
Library student staff	10	30		
Selection & creation of digital images			6,200	7,440
Digitization of images			2,800	3,360
Web site design			1,500	1,500
Preparation of images for class use				
Library student staff (mounting photos, etc.)	310	930		
Teaching Fellows (selecting photos)	980	2,940		
Teaching Fellows (selecting slides, 56 hrs)	1,120	3,360	1,120	3,360
Preservation of images				
Library student staff	45	271		
Collection shelving space (capital)	70	417		
Collection shelving space (maintenance)	19	113		
Digital storage and access			470	2,049
Study space				
Photo study gallery (capital)	2,986	8,959		
Photo study gallery (maintenance)	812	2,436		
Totals	$7,149	$21,849	$18,290	$25,149
Film/photo less digital			($11,141)	($3,300)
Productive (unproductive) use of resources				-13%
Funding source				
Library budget	1,163	3,624	17,170	21,789
Art history department	2,100	6,300	1,120	3,360
University space costs	3,887	11,925	0	0
Totals	$7,149	$21,849	$18,290	$25,149

Before the Scully Project, the university incurred about $7,000 in academic support costs for Professor Scully's course in the year it was taught. These costs over a six year period, during which the course would be taught three times, are estimated at $22,000. As deployed in the Fall of 1996, Web-site support for Professor Scully's course cost an estimated $18,000, or $25,000 over a six-year period. The result is a $3,000 balance arguing against digital provision of images in Professor Scully's course, or a 13% productivity loss in the use of university resources. However, a longer amortization period clearly works in favor of digital provision. The cost model suggests that the break even point on the productive use of university resources comes in eight rather than six years.[19] This happens because:

- The higher absolute cost of the digital images results from one-time staff and vendor cost of converting analog images to digital format. While there is little incremental growth in these costs over six years, staff costs for providing analog images grows linearly. The long-term structure of these costs favors digital provision.
- The cost of the "real" space of bricks and mortar needed to house the photo collection is substantial and grows every year. Similarly, the operation and maintenance of physical space carries the relative high increases of costs for staff and energy. By contrast, the "virtual" space of digital media is relatively inexpensive to begin with, and its unit cost is falling rapidly. Again, the long-term

BEST COPY AVAILABLE

structure of costs favors digital provision.

- To secure the cost benefits of digital provision, an institution would need to increase the operating budget of its library while it reduced spending on Teaching Fellows and space (see the summary display of Funding Sources). More generally, an institution would need to manage its operating and capital budgets as, in significant measure, fungible. The commonplace failure to do this in higher education deprives us of important opportunities to increase institutional productivity.

Along with the amortization period, the number of images digitized is a another major variable that can be used to lower the total cost of digital provision and so move toward a productive use of resources. For years, it has been possible to mount no more than 400 photos in the study gallery. As Table 2 shows, if the Scully Web site had contained 400 digital images, rather than 1,250, conversion costs (italicized to isolate the changes from Table 1) would drop significantly and the six year cost of digital provision ($11,500) would be significantly under the cost of analog provision ($22,000). There is a $10,000 balance in just six years favoring digital provision, or a 88% increase in the productive use of resources.

TABLE 2, "WHAT IF" CONDITION #1: 400 images used primarily for memory training

1st Year and Cumulative 6-Yr Expenses	**400 Photos**		**400 Digital Images**	
Selection of images	**1st Year**	**6-Year Total**	**1st Year**	**6-Year Total**
Full-time library staff for photo collection	797	2,392	*2,067*	*2,480*
Library student staff	10	30		
Selection & creation of digital images			*2,067*	*2,480*
Digitization of images			*933*	*1,120*
Web site design			1,500	1,500
Preparation of images for class use				
Library student staff (mounting photos, etc.)	310	930		
Teaching Fellows (selecting photos)	980	2,940		
Teaching Fellows (selecting slides, 56 hrs)	1,120	3,360	1,120	3,360
Preservation of images				
Library student staff	45	271		
Collection shelving space (capital)	70	417		
Collection shelving space (maintenance)	19	113		
Digital storage and access			157	682
Study space				
Photo study gallery (capital)	2,986	8,959		
Photo study gallery (maintenance)	812	2,436		
Totals	$7,149	$21,849	$7,843	$11,622
Film/photo less digital			($694)	$10,227
Productive (unproductive) use of resources				88%
Funding source				
Library budget	1,163	3,624	6,723	8,262
Art history department	2,100	6,300	1,120	3,360
University space costs	3,887	11,925	0	0
Totals	$7,149	$21,849	$7,843	$11,622

The choice between 400 and 1,250 images has a dramatic impact on costs and productivity. That being so, one must ask what motivates the choice and what impact it has on student learning. Further consideration of this "what if" case is best deferred to the discussion of student productivity .

Speculation about another "what if" case is worthwhile. Professor Scully and his Teaching

Fellows made no use of the Web site in the lecture hall or discussion sessions.[20] What if they had been able to depend on it, instead of traditional slides, for their face-to-face teaching? There is of course a warm debate on whether digital images can match film images in quality or ease of classroom use. The question posed here speculatively assumes no technological reason to favor either analog or digital media, and focuses solely on what happens to costs when classroom teaching is factored in.

Two changes are identified (in italics) in Table 3. They are the cost saving when Teaching Fellows no longer need to assemble slides for the three classroom discussion sessions each conducts during the term and the added cost of equipping a classroom for digital instruction.

TABLE 3, "WHAT IF" CONDITION #2: 1,250 images used for memorization and instruction

1st Year and Cumulative 6-Yr Expenses	400 Photos		1,250 Digital Images	
Selection of images	1st Year	6-Year Total	1st Year	6-Year Total
Full-time library staff for photo collection	797	2,392	6,200	7,440
Library student staff	10	30		
Selection & creation of digital images			6,200	7,440
Digitization of images			2,800	3,360
Web site design			1,500	1,500
Preparation of images for class use				
Library student staff (mounting photos, etc.)	310	930		
Teaching Fellows (selecting photos)	980	2,940		
Teaching Fellows (selecting slides, 56 hrs)	1,120	3,360	*0*	*0*
Preservation of images				
Library student staff	45	271		
Collection shelving space (capital)	70	417		
Collection shelving space (maintenance)	19	113		
Digital storage and access			470	2,049
Study space				
Photo study gallery (capital)	2,986	8,959		
Photo study gallery (maintenance)	812	2,436		
Digitally equipped classroom (capital)			*692*	*2,075*
Digitally equipped classroom (maintenance)			*69*	*208*
Totals	$7,149	$21,849	$17,931	$24,071
Film/photo less digital			($10,782)	($2,222)
Productive (unproductive) use of resources				-9%
Funding source				
Library budget	1,163	3,624	17,170	21,789
Art history department	2,100	6,300	0	0
University space costs	3,887	11,925	761	2,283
Totals	$7,149	$21,849	$17,931	$24,071

This "what if" modeling of the Scully Project shows a $2,000 negative balance, or a 9% loss in productivity. While digital provision in this scenario is not productive within six years, the significant comparison is with the 13% loss in productivity without using digital images in the classroom (Table 1). The conclusion is that substituting digital technology for the labor of selecting slides is itself productive and moves the overall results of digital provision toward a productive use of university resources. This conclusion is strongly reinforced if one considers a variant "what if" condition, in which the Teaching Fellows teach not just three of these discussion sessions in a classroom but all fourteen of them, and where each Fellow selects his or her own slides instead of depending in considerable measure on slides selected by the head Teaching Fellow. This scenario is modeled in Table 4.

TABLE 4, "WHAT IF" CONDITION #3: 1,250 images used for memorization and instruction

1st Year and Cumulative 6-Yr Expenses	400 Photos		1,250 Digital Images	
Selection of images	1st Year	6-Year Total	1st Year	6-Year Total
Full-time library staff for photo collection	797	2,392	6,200	7,440
Library student staff	10	30		
Selection & creation of digital images			6,200	7,440
Digitization of images			2,800	3,360
Web site design			1,500	1,500
Preparation of images for class use				
Library student staff (mounting photos, etc.)	310	930		
Teaching Fellows (selecting photos)	980	2,940		
Teaching Fellows (selecting slides, 700 hrs)	14,000	42,000	*0*	*0*
Preservation of images				
Library student staff	45	271		
Collection shelving space (capital)	70	417		
Collection shelving space (maintenance)	19	113		
Digital storage and access			470	2,049
Study space				
Photo study gallery (capital)	2,986	8,959		
Photo study gallery (maintenance)	812	2,436		
Digitally equipped classroom (capital)			*3,358*	*10,075*
Digitally equipped classroom (maintenance)			*336*	*1,008*
Totals	$20,029	$60,489	$20,864	$32,871
Film/photo less digital			($835)	$27,618
Productive (unproductive) use of resources				84%
Funding source				
Library budget	1,163	3,624	17,170	21,789
Art history department	14,980	44,940	0	0
University space costs	3,887	11,925	3,694	11,083
Totals	$20,029	$60,489	$20,864	$32,871

As a comparison of Tables 3 and 4 indicates, the weekly cost of selecting slides in this new scenario increases twelve-fold, while the use of the electronic classroom increases five-fold. That the classroom costs are absolutely the lower number to begin with also helps drive this scenario to the highly favorable result of an 84% increase in productivity.

In considering these scenarios, it is important to emphasize they all assume funds for Teaching Fellows are fungible in the same way that the library's operating and capital budgets are assumed to be fungible. Faculty and graduate students are most unlikely to make that assumption. Graduate education is one of the core products of a research university. The funds that support it will not be traded about in the way one imagines trades between the operating and capital funds being made for a unit, like the library, that supports education but does not constitute its core product.

Productivity gains subject to reader control

Having accounted for the costs and potential productivity gains that are substantially under the university's administrative control, I will look briefly at potential productivity gains that lie beyond such control--the productivity of readers. In doing this we must consider the value of the qualitative differences between film and digital technologies for supporting Professor Scully's course. The availability of the images throughout the semester at all times of day and night, rather than just before exams, and the large increase in the number of images available for study constitute improvements in quality that make any discussion of increased productivity difficult--but interesting and important as well.

Students were enthusiastic about the convenience of the Web site. They could examine the images more closely, without competing for limited viewing space, at any time they wished. Without question this made their study time more efficient and possibly--though the evidence is inconclusive--more effective.

Let us focus first on the possibility that, as one of the Teaching Fellows observed, students learned more easily but did not learn more. Let us imagine, arbitrarily, that on average students were able to spend two hours less on memory training over the course of the semester because of easy and effective access to digital images. What is the value of this productivity gain for each of Professor Scully's 500 students? It would probably be possible to develop a dollar value for it, related to the direct cost and the short-term opportunity cost of attending Yale. Otherwise, there is no obvious way to answer the question, because each student will appropriately treat the time as a trivial consideration and use it with no regard for the resources needed to provide it. Whether the time is used for having coffee with friends, for sleeping, for volunteer community work, for additional study and a better term paper, or in some other way, the student alone will decide about the productive use of this time. And because there is no administrative means to cumulate the time saved or bring the student's increased productivity to bear on the creation of the information systems that enable the increase, there is no way to use the values created for the student in the calculation of how productive it was to spend library resources on creating the Scully Project.

The possibility that students would use the time they gain to prepare better for tests or to write a better paper raises the issue of quality improvements. How are we to think about the possibility that the teaching and learning libraries support with digital information might become not only more efficient and productive, but also just better? What are the measures of better, and how were better educational results actually achieved? Was it, for instance, better to have 1,250 images for study rather than 400? The head Teaching Fellow answered with an unequivocal yes, affirming that she saw richer, more thoughtful comparisons among objects being made in student papers. But some student responses suggested they wanted to have on the Web site only those images they were directly responsible for memorizing--many fewer than 1,250. Do more images create new burdens or new opportunities for learning? Which objectives and what standards should guide decisions about enhancing instructional support? In the absence of some economically viable way to support additional costs, how does one decide on quality enhancements?

Such questions about quality traditionally mark the boundary of productivity studies. Considerations of quality drive us to acknowledge that, for education, we generally do not have the two essential features needed to measure productivity: clear measures of outputs and a well-understood production technology that allows one to convert inputs into outputs.[21] In such an environment, we have generally avoided talking about productivity for fear that doing so would distort goals--as when competency-based evaluation produces students who only take tests well.[22] Moreover, the rhetoric of productivity can undermine socially rather than empirically validated beliefs among students, parents, and the public about how higher education achieves its purposes. All institutions of higher education depend fundamentally on the maintenance of such socially-validated beliefs.

So I end this account of the Scully Project by observing that what we actually did was marginally not productive, but could readily be made so by extending the amortization period for the Project, or by reducing the number of images provided to students.[23] It also appears that the Project made study much more convenient for students and may well have enhanced their learning. Such quality improvement, even without measurable productivity gain, is one of the fundamental objectives of the library.

These are conditionally positive findings about the economic productivity and educational value of a shift from photographs to digital images to support instruction in the history of art. Such findings should be tested in other courses and, if confirmed, should guide further investment in digital imaging. The soft finding that the use of digital images in the classroom may be productive is heartening, given that digital images may support improvements in the quality of teaching by simplifying the probing of image details and by enabling much more spontaneity in classroom instruction.[24]

All of my arguments about the Scully Project posit that new investment in digital technology would be supported by reduced spending elsewhere. However, doing this would be difficult, forcing us to regard capital and operating budgets--especially the funds that support both "real" and "virtual" space--as fungible. Other possible cost shifts might involve even more fundamental difficulties. It is, for instance, a degree requirement at Yale that graduate students in the History of Art participate in undergraduate instruction. Teaching discussion sections in Professor Scully's course is often the first opportunity graduate students take for meeting this academic requirement. For this reason and others, none of the shifts imagined in the scenarios described above would be easily achieved, and some would challenge us to revisit strongly embedded administrative practices and academic values. Funds rarely flow across such organizational boundaries. Failing to make at least some of these shifts would, however, imperil our ability to improve the quality and productivity of higher education.

PRODUCTIVITY AS AN URGENT CONCERN OF HIGHER EDUCATION

For a long time, higher education has behaved as if compelling opportunities for improving student learning should be pursued without much attention to productivity issues. Our community has focused on desirable results, on the outputs of the productivity formula, without disciplined attention to the inputs part of the equation.[25] One result has been that expenditures per student at public universities in the United States grew between 1979 and 1989 at an average annual rate of 1.82% above inflation. The annual growth rate for private universities was a much higher 3.36%.[26]

It is hard to believe such patterns of cost increase can be sustained much longer or that we can continue simply to increase the price of higher education as the principal means for improving it, and especially for meeting apparently insatiable demands for information technology. We must seriously engage with issues of productivity. Otherwise, there will be little to determine the pace of technology innovation except the squeaky wheel of student or faculty demand or, less commonly, an institutional vision for technology-enhanced education. In neither case is there economically cogent guidance for the right level of investment in information technology. We are left to invest as much as we can, with nothing but socially-validated political and educational ideas about what the phrase "as much as we can" actually means. Because we so rarely close the economic loop between the productivity value we create for users and our investment in technology, the language for decision making almost never reaches beyond that of improving convenience and enhancing quality. I believe it is vitally important for managers of services must become much more productive.[30] Arguments about the incompatibility of higher productivity and the maintenance of quality care resonate strongly with parallel arguments about the impossibility of making higher education more productive without compromising quality. What makes the health care debate so instructive is that we already know which side will prevail. Everywhere we turn, medical institutions and the practitioners who lead them are scrambling to find ways to survive within a managed care environment. Survival means the preservation of quality care, to be sure, but the ineluctable reality is that quality will now be defined within terms set by managed care. We will find ways to talk about increased productivity and quality as complementary rather than as antithetical ideas.

Given the current state of public opinion about higher education, it is impossible for me to believe that we will not soon follow health care. We will almost certainly find ourselves embroiled in divisive, rancorous debates about higher education reform. I hope we will avail ourselves in these debates of a language about information technology that continues to embrace ideas of convenience but reaches strongly beyond them. We will need to talk meaningfully about productivity and link our ability to create productivity gains with investment in information technology. And I hope we will follow the medical community in working to make productivity and quality regularly cognate rather than always antagonistic ideas.

For the last 150 years or so, libraries have been the guardians in the Western world of socially equitable access to information. That is what it has meant for libraries to become public institutions, instead of institutions serving powerful elites, as they once were. This is a noble heritage and a worthy ongoing mission for our profession. And information technology will play a key role in advancing it. As Richard Lanham argues in a landmark essay, "if our business is general literacy, as some of us think, then electronic instructional systems offer the only hope for the radically leveraged mass instruction the problems of general literacy pose."[31] But unless information technologies are employed productively, they will not offer the leverage on information access and literacy that Lanham and others of us hope for. Indeed, unless those of us who manage libraries and other instruments of scholarly discourse are prepared to embrace the language of productivity, we will find our ability to provide socially equitable access to information weakened as decisions are made about where investments for democratic education will be directed. I look at managed health care and the Western Governors' University and fear that traditional universities and their libraries will lose ground, not because we have failed to embrace information technology, but because we have failed to embrace it productively. I fear that outcome most because it imperils the wonderful accomplishment of libraries and because it could significantly weaken the public good that free libraries have been creating for the last 150 years.

Scott Bennett
University Librarian
Yale University

information technology to understand the fundamental economic disconnect in the language of convenience and service we primarily use and to add the language of productivity to our deliberations about investing in information technology.

In connecting productivity gains with technology investment, we may find--as analysis of the Scully Project suggests--that some improvements can be justified while others cannot. Productivity measures should not be the sole guide to investment in information technology. But by insisting on securing productivity gains where we can, we will at least identify appropriate if sometimes only partial sources for funding new investments and thereby lower the rate at which overall costs rise in higher education above those in the rest of the economy.[27]

The stakes for higher education in acting on the productivity problems confronting it are immense. Today, it is regularly asserted that administrative activities are wasteful and should be made more productive. But turning to core academic activities, especially teaching, we feel that no productivity gains can be made without compromising quality. Teaching is rather like playing a string quartet. It required four musicians in Mozart's day, and it still does. To talk about making the performance of a string quartet more productive is to talk patent nonsense. To talk about making classroom teaching more productive seems to many almost as objectionable. The observable result is that higher education has had to live off the productivity gains of other sectors of the economy. The extreme pressure on all of higher education's income sources suggests we are coming to the end of the time when people are willing uncritically to transfer wealth to higher education. Socially validated beliefs about the effectiveness of higher education are in serious jeopardy.[28] If our community continues to stare blindly at these facts, if we refuse to engage seriously with productivity issues on an institutional and community-wide basis, we will bring disaster upon the enterprise of teaching and learning to which we have devoted our professional lives.

If this seems alarmist, consider the work of ten governors in the western United States intent on creating a high-tech, virtual university, the Western Governors' University.[29] Faced with growing populations and burgeoning demand for higher education, but strong taxpayer resistance to meeting that demand through the traditional cost structures of higher education, state officials are determined to create a much more productive regional system of higher education. That productivity is the key issue is evident in the statement of Alvin Meiklejohn, the chairman of the State Senate Education Committee in Colorado. "Many students in Colorado," he said, "are now taking six years to get an A.B. degree. If we could reduce that by just one year . . . it would reduce the cost to the student by one-sixth and also free up some seats in the classrooms for the tidal wave we see coming our way" (*New York Times*, 25 Sept. 1996, p. B9). Senator Meiklejohn is looking for a 17% increase in productivity. I think library and information technology managers know where some of that gain may be found. If however we scoff at the idea of increasing student productivity through the use of information technologies, if we insist that the job of measuring and redirecting the productivity gains we create with information technology is impossible, if we trap ourselves in the language of convenience and fail to engage with issues of productivity, then the consequences--at least in the West--are clear. Major new investment in higher education will be directed not to established institutions but to new organizations that can meet the productivity standards insisted on by Senator Meiklejohn and the taxpayers he represents.

A second and larger ground swell in American life is also instructive on the question of productivity. Health care reform and managed care are both driven by the idea that the high cost and poor delivery of health care must change, that costs must be controlled--that health care

APPENDIX: COST MODEL FOR THE SCULLY PROJECT

The cost model uses the following facts, estimates, and assumptions:

Introduction to the History of Art, 112a
- Course offered once every two years; three times in six years
- Number of students enrolled in Scully course = 500/term
- Number of weeks Scully photos available in study space = 9 weeks per term
- Length of term = 14 weeks
- Number of Teaching Fellows for Scully course = 25
- Approximate value/hour of Teaching Fellow time = $20
- Hourly wage for library student staff = $6.46

Staff costs for selection, maintenance, and display of slide & photo images
- 1 FTE permanent staff devoted to photo collection = $xx,xxx for salary and benefits
- % of permanent library staff effort devoted to Scully course = x%
- Library student staff devoted to photo collection = 40% of $11,500 = $4,600 at $6.46/hr = 712 hrs
- Library student staff devoted to exhibiting Scully photos = 48 hrs/year
- Time spent by Teaching Fellows assembling photo study =3.5hr/wk*14 wks = 49hrs
- Time spent by Teaching Fellows assembling slides for review classes = 56hrs

Cost to prepare digital images for instructional use
- Number of images in Scully Project = 1,250
- Digitization of images (outsourced) = $2,800
- Change in Scully Project Web site content over 6 years = 20%
- Selection and creation of images (by 2 Teaching Fellows) = $6,200
- Web site design = $1,500

Preservation and access costs for slide, photo, and digital images
- Library student staff hours spent on mending & maintenance of photos = 7 hrs/year
- Disk space required for Scully Project = .855 GB
- Disk space required per volume for Project Open Book = .015 GB
- Scully Project images = 57 Open Book vols
- Digital Storage costs = $2.58/year/Open Book vol.
- Digital access costs = $5.67/year/Open Book vol.
- Storage and access cost inflation = -13%/year

Study and other space costs
- Number of items in photo collection = 182,432
- Number of Scully photos mounted in study space = 200 for mid-term; 400 for final
- NSF of photo collection in Street Hall = 1,733
- NSF collection shelving for Scully photos =400/182,432 * (1,733-500)= 2.7
- NSF of photo study space = 2019 + .25*1500=2,394
- % of photo study space devoted to Scully photos per term = 20%
- NSF of photo study space available for Scully photos = 2,394 * .2 * (9/28) = 154
- NSF of photo study space utilized during term = 154 * 75% = 116
- Annual cost of space maintenance = $7 NSF
- Cost of new construction = $300 NSF
- Amortization of capital costs at 8% over 35 yrs = $85.81 per $1,000
- Capital cost of converting existing classroom for digital display = $50,000 depreciated over 6 years
- Maintenance of digital classroom hardware and software = 10% of capital cost/year = $5,000/year
- Availability of digital classroom = 8 class hours*5 days/wk*28wks*.8 efficiency factor = 896 sessions/yr
- Need by Scully grad. assistants for digital classroom sessions = 25*3= 75 sessions/yr = 8.3% of avail. sessions

ENDNOTES

……

For additional information about the conference, or The Andrew W. Mellon Foundation's scholarly communication initiatives, please contact Richard Ekman. For additional information about ARL or this web site contact Patricia Brennan, ARL Program Officer at (202) 296-2296.

Return to Office of Scholarly Communication Home Page

ARL Home ARL Scholarly Communication and Technology Home Page

© Association of Research Libraries, Washington, DC
Web Design by Angelo F. Cruz
Maintained by ARL Web Administrator
Last Modified: August 29, 1997

Scholarly Communication and Technology

Conference Organized by The Andrew W. Mellon Foundation

at Emory University
April 24-25, 1997

Copyright © of the papers on this site are held by the individual authors or The Andrew W. Mellon Foundation. Permission is granted to reproduce and distribute copies of these works for nonprofit educational or library purposes, provided that the author, source, and copyright notice are included on each copy. For commercial use, please contact **Richard Ekman** at the The Andrew W. Mellon Foundation.

Session #8 Sustaining Change

Cost and Value in Electronic Publishing

James J. O'Donnell
Professor of Classical Studies and Vice Provost (Interim)
Information Systems and Computing
University of PennsylvaniaYale University

Gentlemen, Here is the current version of my paper from the conference. It follows closely what you had seen before with the main amendment of inserting the illustrative bit about indulgences, which arose in situ in response to something naive Andy Odlyzko had said and seemed usefully illuminating. This is not the hard-edged analytical stuff that gave that conference its best moments, but it may still have its use. I'm very open to any editorial suggestion, etc., that you may have. If you would like it in some other electronic form or even on, gasp, paper, I'd be happy to supply that as well.

James J. O'Donnell
June 24, 1997
via e-mail

Cost and Value in Electronic Publishing
J.J. O'Donnell

This paper is read perhaps best through binocular lenses. On the one hand, it is an account of the value and function today and for the foreseeable future of the kinds of electronic networked texts. But on the other hand, it questions our ability to account for such value and function. In search of the particular, it risks the anecdotal; in defense of value, it expresses skepticism about calculations of cost and price.

I am a student of the works of St. Augustine and shall begin accordingly with confession. The single most transforming feature of cyberspace as we inhabit it in 1997 for my own scholarship can be found in a warehouse on the edges of downtown Seattle. I mean the nerve center of www.amazon.com. I have conducted strikingly extensive experiments over the last year and can now say conclusively that it is possible to go from a supine position on my living room sofa, just vaguely tickled by the thought of a book I might be interested in, to a seated position a few feet away in my study striking the "Return" key to complete and execute an order for the book, which will appear 48-72 hours later at my office, in three minutes flat. The impact, retrospective and prospective, on the finances of my sector of higher education, could well be catastrophic. Participants in this conference will immediately recognize that I speak not merely of the cost of the books and the cost of my time reading, or my time feeling guilty about not reading, them, but also of course the cost of space on my shelves and the cost of my time and energy reshelving them each time I take them down to read, or to feel guilty about not reading, them. A couple of months ago, I had the chance to take the tour of Amazon.com's facilities vigor and excitement that positively swirls over the printed word as electronic media of communication are used to whisk volumes to all parts of the world.

If my approach seems whimsical, do not be misled. The real habits of working scholars often fall outside the scope of discussion when new and old forms of publication are considered. I will have some things to say shortly about the concrete results of surveys we have done for the Bryn Mawr Reviews project funded by Mellon, and more of our data appear in the paper by my colleague Richard Hamilton, but I want to emphasize a few points by personalizing them first.

First, and most important, Amazon books is a perfect hybrid: a cyberspace service that delivers the old technology better and faster than ever before. As such it may seem to be no more than an exemplification of the old McLuhan dictum that I like to quote, that the content of a new medium is an old medium. But we need to pay closer attention to what happens to books when they begin to move faster and in greater quantities.

Second, therefore, my ritual allusion to the paradox of the scholar wallowing in information that he does not actually read is not merely humorous: it is a fact of life. The file drawers full of photocopies, read and unread, that every working humanist seems now to possess are a very recent innovation. As best I can recall for myself, they started to accrue around 1980, toward the end of my time as an assistant professor. When the joking began -- "Once you photocopy the article, you don't have to read it" -- I cannot say, but I suggest it marks an important self-awareness. Photocopying is a service that has declined sharply in price -- if measured in real terms -- over the last twenty years, and it is certainly the case that graduate and undergraduate students can tell the same joke on themselves today. Perhaps only full professors today reach the point where they can joke similarly about books, but if so surely we are the leading edge of a wedge. The "superstores" brought scholarly bookbuying to more eyes and fingertips than ever, starting about five years ago, and now on-line sales offer the opportunity more broadly. It is very certainly the case, for example, that the city where I went to high school, the nineteenth largest in population in the US today, was still in the summer of 1995 when last I visited it, exactly the desolate wasteland for book purchasers that it was when I haunted a few miserable shops 30 years ago (cherishing the small rack of distinctively covered Scribner paperbacks, for example). But in the two years since, it has acquired a Barnes and Noble superstore and at the same time anyone with Internet access is now just as close to that Seattle warehouse as I am. They joke that they run the world's largest bookstore, with 42 million locations around the world. The joke has a point to it. (Among other things, 30% of Amazon's business is already overseas. It makes perfect sense to think that a mechanism for speeding delivery of American books would be well-received abroad.)

But abundance is not wealth, for wealth is related to scarcity. This, I think, is the point of our jokes. When each new book, pounced on with delight in a bookstore, was an adventure, and when each scholarly article was either a commitment of time or it was nothing, the mechanical systems of rationing that kept information scarce also kept it valuable. But if we now approach a moment when even quite serious books are abundantly available, then their individual value will surely decline. To continue in confessional vein a moment, I think I have seen this when moving house a couple of times in the last couple of years. Dignified, serviceable, but somewhat tired hard-cover copies of well-regarded fiction -- George Eliot, say, or Henry James -- the sort of thing I used to snatch up with pleasure for $2 in a second-hand shop, to lay by against the time when I would read them: these veterans, whether read or not, have found themselves heading back to the second-hand shops. Not because my respect for the texts, or my guilt at not *yet* having read them, is any the less, but because I know that when I find I really do need to read *Daniel Deronda* --- a need I am quite sure will arise someday -- I have come to be confident that there will be a superstore, or an Internet terminal, close to hand. Eliot hasn't yet declined in value, but I am content to point out that our calculations of such value are made on a slippery slope.

(I am fond of historical illustration. A student of mine at Penn is now working hard on a dissertation that involves late medieval indulgences -- not just the theological practice of handing out remission of punishment but the material media through which that remission was attested. It turns out there were indeed some very carefully-produced written indulgences before printing was introduced, but indulgences were among the first printed artifacts ever. The sixteenth century saw a boom in the indulgence business as mass-production made the physical testimony easier to distribute and obtain. The "information economy" of indulgences showed a steady rise through several generations. [The *price* history of indulgences seems still obscure, for reasons my student has not yet been able to fathom; it would be-interesting to see if supply and demand had more to do with the availability of the artifact or was rather measured by the number of years or purgatorial remission.] But there came a point at which, almost at a stroke, the superabundance of printed indulgences was countered by loud assertions of the worthlessness of the thing now overpriced and oversold. There followed the familiar cycle of business process re-engineering in the indulgence business: collapse of market, restructuring, downsizing, and a focusing on core competencies. The indulgence business has never been the same.)

A third and last confessional point. As founding co-editor of Bryn Mawr Classical Review (BMCR) since 1990, I think I may reasonably assert that I have been thinking about and anticipating the benefits of networked electronic communication for scholars for some time now. Yet, as I observe my own practices, I must accept that my powers of prognostication have been at best imprecisely focused. Yes, a network connection at my desktop has transformed the way I work, but it has done so less through formal deployment of weighty scholarly resources and more through humbler tools. I will list a few:

1. On-line reference: Though I happened to have owned the compact OED for over twenty years and now in fact own a set of the Encyclopedia Britannica, I rarely in fact used the former and rarely remember to look at the latter. But their electronic avatars I consult now daily: "information" sources on myriad topics far more detailed and scholarly than any previously in regular use. This process went so far that in 1994, I found myself giving away my compact (magnifying-glass edition) OED as simply too bulky and not enough useful beside the electronic version. On the other hand, my trusted, not to say revered, desk copy of Henry Fowler's *Concise Oxford Dictionary* sees hardly any use at all: I consult the more comprehensive resource for ready reference. (Greg Crane of Tufts University reports that the same phenomenon has occurred with the various on-line versions of the standard Liddell-Scott lexicon of Greek literature that he has created. Though the concise desk dictionary is available, users regularly and overwhelmingly prefer the "unabridged" version.)
2. On-line productivity information: Under this category I include far better information about weather and travel weather than ever before; access to current airline schedules and other travel information including hotel directories; nationwide telephone directories including yellow pages; on-line newspapers and newsfeeds; and -- essential reading for anyone lately gone over from the traditional academic life to managing a large staff -- a daily update of the latest "Dilbert" cartoon. I no longer purchase newspapers (with the interesting effect that I am less well-informed about Philadelphia than I have ever been: my Philadelphia awareness used to come as a bonus along with world and national news either by newspaper or at 11 p.m. on TV, but now my news needs are satisfied without ever having to find out what is going on within blocks of my residence), and my forty-year-long habit, going back to when I learned to read as a child, of consulting the *World Almanac* for every factual question, is fading.
3. E-mail as productivity tool: The positive impact of e-mail communication on scholarship for me cannot be underestimated. Relatively little of my e-mail has to do with my scholarship, but that proportion is important first of all: news of work in progress, often including copies of papers, and ongoing conversation with specialists elsewhere is a great boon, no question. But the real enhancement comes from the way e-mail lets me handle more mundane responsibilities. I have far more contact with my students than ever, and spend much less time sitting in my office for "office hours" waiting for them to turn up. With the staff who now report to me, ordinary business gets done on quick turnaround almost in real time. With both students and staff, face to face time is increasingly used for more substantial interaction and less busy work. There really are fewer meetings.
4. Formal on-line publishing endeavors: I confess that I use the kinds of resources that Mellon grants support far less than I might have expected. I did indeed point my students to a specific article in a MUSE journal a few months ago, and I browse and snoop, but it was only in writing this paper that I had the excellent idea to bookmark on my browser MUSE's Journal of Early Christian Studies and JSTOR's Speculum -- they appear just below the exciting new URL for the New York Times Book Review on-line.

So we, or at least I, live in a world where electronic and print information are already intermarrying regularly, where the traditional content of print culture is declining in value, and where the value of electronic information is not so much in the content as in the interconnectedness and the greater usefulness it possesses. For a conference as explicitly devoted as this one is to carrying traditional resources into electronic form, all three of those observations from experience should give pause. In fact, I am going to argue that the intermediacy and incompleteness of the mixed environment we inhabit *today* is an important and likely *durable* consideration. We must be careful not to imagine ourselves forward too quickly into a transformed and perfected world that we may in fact never reach. The implications of this argument will return later in this paper. To give them some weight, let me recount and discuss some of our experiences with BMCR. For some in this audience, there will be some familiar tales told here, but with I hope fresh and renewed point.

****

When we began BMCR, we wrote around to publishers with classics lists and asked for free books. An engaging number responded affirmatively, considering we had no track record. Oxford Press sent many books, Cambridge Press did not respond: a 50% success rate with the most important British publishers seemed very satisfactory for a startup. During our first year, we reviewed many OUP books, few if any Cambridge titles. There then appeared, sometime in 1991 or 1992, an OUP Classics catalogue, with no fewer than two dozen titles appending blurbs from "Bryn Mawr Classical Review." (From this we should draw first the lesson that brand names continue to have value: OUP could have chosen to identify its blurbs, as it more commonly does, by author of the review than by title of the journal, but we had chosen our "brand" well.) Approximately two weeks after the OUP catalogue appeared, we received unsolicited a first handsome box of books from Cambridge, and we now have a happy and productive relationship with both publishers. Our distinctive value to publishers is our timeliness: books reviewed in time to blurb them in a catalogue while the books are still in their prime selling life, not years later. The practical value to scholars is that information about and discussion of current work moves more rapidly into circulation. (Can a dollar price be placed on such value? I doubt it. I will return later to my belief that one very great difficulty in managing technology transitions affecting research and teaching is that our economic understanding of traditional practices is often too poor and imprecise to furnish a basis for proper analysis. In this particular case, we must cope with the possibility that a short-term advantage will in the long term devalue the information by increasing its speed of movement and decreasing its lifetime of value.)

We began BMCR in part because we had already in place a circle of collaborators. Rick Hamilton had created Bryn Mawr Commentaries in 1980, offering cheap, serviceable, reliable texts of Greek and Latin authors with annotation designed to help real American students of our own time; in a market dominated by reprints of texts for students in the upper forms of British public schools in another century, the series was an immediate hit. It quickly became the most successful textbook series in American classics teaching. I had joined that project in 1984 and in slightly over a decade we had almost 100 titles in print. In the course of that project, Hamilton had assembled a team of younger scholars of proven ability to do good work on a short deadline without exclusive regard for how it would look on a c.v. -- textbook-writing is notoriously problematic for tenure committees. This group formed the core of both our editorial board and our reviewing team. If you had asked us in 1990 what we were doing, we would have said that we were getting our friends to review books for us. This was true insofar as it meant that we could do a better job more quickly of getting good reviews moving because we had already done the work of building the community on which to draw.

But what surprised us most was that a little more than a year after we began work, we looked at the list of people who had reviewed for us and found that it had grown rapidly beyond the circle of our friends and even the friends of our friends. A book review journal seems unusually well situated to build community in this way, because it does not wait for contributions: it solicits them and even offers small compensation -- free books -- to win people over. If then it can offer timely publication, at least in this field, it is possible to persuade even eminent and computer-hostile contributors to participate. (To be sure, there are no truly computer-hostile contributors left. The most recent review we have published by someone not using at least a word processor is three years old.)

But the fact of networked communication meant that the reviewer base could grow in another way. A large part of our working practice, quite apart from our means of publication, has been facilitated by the Internet. Even if we only printed and bound our product, what we do would not be possible without the productivity-enhancement of e-mail and word processing. We virtually never "typeset" or "keyboard" texts, a great saving at the outset. But we also do a very high proportion of our communication with reviewers by e-mail. Given the difficulties of moving formatted files across platforms that persist even now, we still receive many reviews on floppy disks with accompanying paper copies to assure accuracy, but that is only a last step in a process greatly speeded by the speed of optical fiber.

Further, in July 1993 our imitation of an old practice led to a fresh transformation of our reviewing population. We began to publish a listing of "books received" -- enough were coming to hand to make this seem like a reasonable practice, one we now follow every month. By stroke of simple intuition and good luck, Hamilton had the idea to prepend to that list a request for volunteers to review titles yet unplaced. (I may interpose here that Hamilton and I both felt acutely guilty in the early years every time one or two books were left after several months unplaced for review. Only when we read some time later the musings of a book review editor for a distinguished journal in another field well known for its reviews and found that he was publishing reviews of approximately 5% of the titles that came to his desk did we start to think that our own practice [reviewing, on a conservative estimate, 60-70% of titles] was satisfactory.) The request for volunteers drew an unexpected flood of requests. We have now institutionalized that practice to the point that each month's publication of the "books received" list needs to be coordinated for a time when both Hamilton and I are prepared to handle the incoming flood of requests: 30-40 a month for a dozen or so still-available titles.

But the result of this infusion of talent has been an extraordinary broadening of our talent pool. Though a few reviewers (no more than half a dozen) are household names to our readers as authors of more than a dozen reviews over the seven years of our life, we are delighted to discover that we have published, in the classical review journal alone, 430 different authors from a total of about 1000 reviews. Our contributors come from several continents: North America, Europe, Africa, Asia, and Australia. By the luck of our having begun with a strategy based in praxis rather than ideology (beginning, that is, with people who had contributed to our textbook series), we have succeeded in creating a conversation that ranges widely across disciplinary and ideological boundaries. The difficulty of establishing working relations with European publishers remains an obstacle that perplexes us: but that difficulty chiefly resides in the old technology of postal delays and the fact that even e-mail does not eradicate the unfamiliarity that inheres when too few opportunities for face-to-face encounter exist.

Our experience with Bryn Mawr Medieval Review has been instructively different. There we began not with a cadre of people and an idea, but merely with an idea. Two senior editors, including myself, recruited a managing editor who tried to do in a vacuum what Hamilton and I had done with the considerably greater resources described above. It never got off the ground. We put together an editorial board consisting of smart people, but people who had no track record of doing good work in a timely way *with us*: they never really engaged. There was no cadre of prospective reviewers to begin with, and so we built painstakingly slowly. In the circumstances, there was little feedback in the form of good reviews and a buzz of conversation about them, and publication never exceeded a trickle.

We have speculated that there are some intrinsic differences between "classics" and "medieval studies" as organized fields in this country that are relevant here. Classicists tend to self-identify with the profession as a whole and to know and care about materials well beyond their immediate ken. A professor of Greek history can typically tell you in a moment who the leading people in a subfield Latin literature are, and even who some of the rising talent would be. But a medievalist typically self-identifies with a disciplinary field (like "history") at least as strongly as with "medieval studies", and the historian of Merovingian Gaul neither knows nor cares what is going on in Provencal literature studies. I am disinclined to emphasize such disparities, but they need to be kept in mind for what follows.

After two and a half years of spinning our wheels, with to be sure a fair number of reviews, but only a fair number and productivity clearly flagging, we made the decision to transfer the review's offices to new management. We were fortunate in gaining agreement from Professor Paul Szarmach of the Medieval Institute of Western Michigan University to give the journal a home and some institutional support. Western Michigan has been the host for a quarter century of the largest come-all-ye in medieval studies in the world, the annual Kalamazoo meetings. Suddenly we had planted the journal at the center of a network of self-identified medievalists. The managing editorship has been taken up by two WMU faculty, Rand Johnson in Classics and Deborah Deliyannis in History, and since they took over the files in spring 1996, the difference has been dramatic. In the last months of 1996, they had the most productive months in the journal's life and on two occasions distributed more reviews in one month than BMCR did. BMCR looks as if it will continue to out produce BMMR over the next twelve months by an appreciable pace, but the gap is narrowing.

Both BMCR and BMMR stand to gain from our Mellon grant. A new interface on the WWW, a mechanism for displaying Greek text in Greek font, enhanced search capabilities, and other features you may well surmise will be added to what is still the plain-ASCII text of our archives which are still, I am either proud or embarrassed to claim, on a gopher server at the University of Virginia Library. When we began our conversations with Richard Ekman and Richard Quandt in 1993, indeed, one chief feature of our imagined future for BMCR was that we would not only continue to invent the journal of the future, but we would put ourselves in the position of packaging what we had done for distribution to others who might wish to emulate the hardy innovation of an electronic journal. About the time we first spoke those words, Mosaic was born; about the time we received notice of funding from the Mellon foundation, Netscape sprang to life. Today the "NewJour" archive based on a list co-moderated by myself and Ann Okerson on which we distribute news of new electronic journals suggests that there have been at least 3500 electronic journals born -- some flourishing, some already vanished. Though BMCR is still one of the grandfathers of the genre (Okerson's 1991 pathbreaking directory of e-journals listed 29 titles including BMCR, and that list was near exhaustive), we are scarcely exemplary: it's getting crowded out here.

But meanwhile, a striking thing has happened. Our users have, with astonishing unanimity, not complained about our retrotech appearance. To be sure, we have always had regrets expressed to us about our Greekless appearance and our habit of reducing French to an accentless state otherwise seen in print chiefly in Molly Bloom's final soliloquy in the French translation of Ulysses. But those complaints have not increased. Format, at a moment when the web is alive with animation, colors, java scripts, and real audio, turns out to be far less importance than we might have guessed. Meanwhile, to be sure, our usage has to some extent plateaued. During the first heady years, I would send regular messages to my co-editors about the boom in our numbers. That boom has never ended, and I am very pleased to say that we have always seen fewer losses than gains to our subscription lists, but we are leveling out. Where Internet usage statistics continue to seek the stratosphere, we saw a "mere" 14% increase in subscriptions between this time twelve months ago and today. (Our paper subscriptions have always remained very consistent and very flat.) It is my impression that we are part of a larger Internet phenomenon that began in 1996, when the supply of sites began to catch up to demand and everyone's hits-per-site rate began to level off.

****

But we are still a success, in strikingly traditional ways. Is what we do worth it? How can we measure that? My difficulty in answering such questions is that in precisely the domain of academic life that feels most like home to me, we have always been astonishingly bad at answering such questions. Tony Grafton and Lisa Jardine, in their important book on Renaissance education From Humanism to the Humanities, make it clear how deeply rooted the cognitive dissonance in our profession is between what we claim and what we do. Any discussion of the productivity of higher education is going to be inflammatory, and any attempt to measure what we do against the standards of contemporary service industries will evoke defenses of a more priestly vision of what we are and what we can be -- in the face of economic pressures that defer little if at all to priesthoods.

But I will also suggest that there is one additional reason why it is premature to begin measuring too closely what we do. Pioneers are entitled to be fools. Busting sod on the prairie was a disastrous mistake for many, a barely sustainable life for many many more (read Wallace Stegner's luminous memoir *Wolfwillow* for chapter and verse), and an adventure rewarding to few. But it was also a necessary stage towards a productive and, I think we would all agree, valuable economy and culture. I suggest that if we do not know how to count and measure what we do now on the western frontier with any certainty, we do already know how to fret about it. We know what the issues are and we know the range of debate.

By contrast, any attempt to measure the value of electronic texts and images or of the communities they facilitate is premature in a hundred ways. We have no common space or ground on which to measure them, for one thing: a thousand or a million experiments are not yet a system. We do not know what scales, what survives, what has value that proves itself to an audience willing to pay to sustain it. We can measure some of the costs, but academic enterprises are appallingly bad at giving fully-loaded costs, inasmuch as faculty time, library resources, and the heat the keeps the fingers of the assistant typing HTML from freezing are either unaccounted for or accounted for far more arbitrarily than is the case for, for example, amazon.com. We can measure some of the benefits, but until there is an audience making intelligent choices about electronic texts and their uses, those measures will be equally arbitrary.

Let me put it this way. Was an automobile a cost-effective purchase in 1915? I know just enough of the early history of telegraphy to surmise, but not enough to prove, that the investment in the first generation of poles and wires -- Ezra Cornell's great invention -- could never possibly have recouped itself to investors, and in fact as with many other "new technologies" of the nineteenth century one important stage in development was the great crash of bankruptcies, mergers, and reorganizations that came at the end of the first generation. "Western Union," in which Cornell was a principal shareholder, was one economic giant to emerge in that way. A similar thing happened to railroads in the late nineteenth century. Such a reading of history suggests that what we really want to ask is not whether we can afford the benefits of electronic texts but whether and how far we can allow universities and other research institutions to afford the risks of such investment.

For we do not know how to predict successes: there are no "leading economic indicators" in cyberspace to help us hedge and lay our bets. Those of us who have responsibility for large institutional ventures at one level or another find this horribly disconcerting, and our temptation over the next months and years is always going to be to ask the tough, green-eyeshade questions, as indeed we must. But at the same time, what we must be working for is an environment in which not every question is pressed to an early answer and in which opportunity and openness are sustained long enough to shape a new space of discourse and community. We are not yet ready for systems thinking about electronic information, for all that we are tempted to it: the pace of change and the shifts of scale are too rapid. The risk is always that we will think we discern the system of the future and so seek to institutionalize it as rapidly as possible, to force a system into existing by closing it off by main force of software, harware, or text-encoding choices. To do so now, I believe, is a mistake.

For one example: "Yahoo" and "Altavista" are powerful tools to help organize cyberspace in 1997. But they are heavily dependent on the relative sizes of the spaces they index for the effectiveness of their results: they cannot in present form scale up. Accordingly, any and all attempts to measure their power and effectiveness are fruitless. For another example: there is as yet no systemic use of information technology in higher education beyond the very pedestrian and pragmatic tools I outlined above. Any attempt to measure one experiment thus falls short of its potential precisely because no such experiment is yet systemic. There is nothing to compare it with, no way to identify the distortions introduced by uniqueness, or by the way the demands of present institutional structures distort an experiment in ways that limit its effectiveness.

What we still lack is any kind of economic model for the most effective use of information technology in education and scholarship: that much must be freely granted. The interest and value of the Mellon grants and this program, I would contend, lies in the curiosity with which various of our enterprises push our camel-like noses under one or another tent flap, in search of rewarding treats. Until we find them, we must, however, be content to recognize that from a distance we all appear as so many back ends of camels showing an uncanny interest in a mysterious tent.

……

For additional information about the conference, or The Andrew W. Mellon Foundation's scholarly communication initiatives, please contact Richard Ekman. For additional information about ARL or this web site contact Patricia Brennan, ARL Program Officer at (202) 296-2296.

Return to Office of Scholarly Communication Home Page

ARL Home ARL Scholarly Communication and Technology Home Page

© Association of Research Libraries, Washington, DC
Web Design by Angelo F. Cruz
Maintained by ARL Web Administrator
Last Modified: August 29, 1997

Scholarly Communication and Technology

Conference Organized by The Andrew W. Mellon Foundation

at Emory University
April 24-25, 1997

Copyright © of the papers on this site are held by the individual authors or The Andrew W. Mellon Foundation. Permission is granted to reproduce and distribute copies of these works for nonprofit educational or library purposes, provided that the author, source, and copyright notice are included on each copy. For commercial use, please contact **Richard Ekman** at the The Andrew W. Mellon Foundation.

Session #9 Summation

Conference Summation Remarks

Deanna B. Marcum
President
Commission on Preservation and Access

* * *

Before I attempt to summarize the elements of the conference that seemed most significant to me, I want to thank Richard Ekman, Richard Quandt, and The Andrew W. Mellon Foundation for having brought this group together. They could have asked all the speakers to submit html versions of their papers, and then made them available on a Web site and created a special listserv to carry on our discussions. But that would have been to lose what was most useful about the past two days -- the opportunity to discuss ideas face to face. Hal Varian, in his after-lunch talk, pointed out that attention is the scarce resource today. By convening this group, the Mellon Foundation allowed us to concentrate our collective attention on the important topic of scholarly communication.

Three of us have been asked to bring our individual perspectives to summarizing a conference that was crowded with excellent speakers. I offer my comments from the perspective of a librarian.

I believe that all the speakers shared at least one major assumption: that the purposes of the library will remain unchanged, though the means through which it achieves those purposes may be quite new and different. The library still exists to provide whatever resources are necessary to meet the research and inquiry needs of students and faculty members. At the same time, the library as a physical place still serves as a community symbol of knowledge and its importance to society.

Against the backdrop of this shared assumption, we heard speakers with at least three different perspectives: 1) technology enthusiasts, who see how technology can change the essential nature of our work and who urge all of us to accelerate the pace of transformation; 2) librarians, who are concerned about managing "hybrid" organizations, which will support massive paper-based collections while also taking full advantage of electronic resources; and 3) publishers, who want to understand how electronic scholarly communication will affect the publishing business.

In all the talks, the speakers eloquently portrayed the promise of technology for increasing access to information. Far less clear were answers to the following questions:

1. Can technology reduce the cost of scholarly communication?
2. Do students learn better when using technology?
3. Are libraries organized to take full advantage of the possibilities for enhanced access?

I found the questions raised by the speakers more compelling than their reports of progress, perhaps because so many of the projects they discussed are not far enough advanced to offer solid conclusions. I would summarize these questions, which came up in many different guises, as follows:

1. Where should we concentrate our efforts -- on converting print documents to digital form to increase access, or on adding digital files that were born digitally to existing library resources? Can we do both?

2. How do we shift the focus from individual institutional holdings to the provision of more extensive access to materials for our students and scholars? How do we budget for this shift?

3. How can digital libraries be discussed without taking into account the networks for delivering information resources and the equipment necessary for reading digital files? Libraries have never been islands unto themselves, but there is increasing awareness of their interdependency.

4. What, exactly, do we want to count? How do we count? Our tradition is to collect quantitative data about the size of collections, budgets, staffs, transactions. If we keep in mind that the library's primary purpose is to provide resources for scholarship and teaching, what should we be counting in the digital environment? Thus far, only one conclusion is clear: counting "hits" on a Web site is useless.

5. Will we be able to read anything we are now producing in electronic form a few years from now? "Digital preservation" has been alluded to many times, but it remains an area of great uncertainty.

The Andrew W. Mellon Foundation has taken an extremely important first step toward helping us understand the implications of scholarly communication in the digital environment by asking project directors to describe in detail the assumptions of their pilot projects and then to be candid about outcomes and users' reactions. This process needs to be continued over time as the projects mature. In the course of their development, I hope that we can learn more about the following areas:

1. **Desirable Future States**

   We heard a great deal about changes we can expect, but we need to have more intense discussions about those changes we are prepared to pursue and effect. Descriptions of the various projects gave us much to ponder. We must now spend more time specifying the desirable future outcomes and conditions against which we can measure project results.

2. **The nature of collections**

   Electronic information resources alter both our notions about the significance of very large collections and our methods of allocating resources for the provision of information. How are these changed perceptions to be accommodated within higher education?

3. **Variations in disciplines**

   There appear to be genuinely different requirements for research resources from discipline to discipline. In describing projects, we should look carefully at the types of resources involved and the audience, or audiences, for them. It is not possible to generalize about what scholars need and want.

4. **Users' views**

   To date, the projects have provided considerable data about how information resources have been scanned and indexed and how they can be retrieved. In the future, we must learn more about users' reactions to the new format and about the utility of digital information to them.

5. **Digital archiving**

   Kevin Guthrie rightly pointed out that there are not technological barriers to archiving and to meeting our societal obligation to preserve the intellectual record. But now we must find the most suitable Ð and the most cost-effective Ð methods for fulfilling that obligation.

Though most of the conference speakers advocated continued support for pilot projects, many also asked that more specific requirements for reporting results be established. All praised The Andrew W. Mellon Foundation for creating an environment of candor and trust for the exchange of sensitive information. The future of scholarly communication may not be clear, but the need for all of us to understand better the implications of electronic publishing is entirely evident. To that process of understanding, this conference was a most valuable contribution.

……

For additional information about the conference, or The Andrew W. Mellon Foundation's scholarly communication initiatives, please contact Richard Ekman. For additional information about ARL or this web site contact Patricia Brennan, ARL Program Officer at (202) 296-2296.

Return to Office of Scholarly Communication Home Page

ARL Home ARL Scholarly Communication and Technology Home Page

© Association of Research Libraries, Washington, DC

Web Design by Angelo F. Cruz
Maintained by ARL Web Administrator
Last Modified: September 4, 1997

468